DOCUMENT RESUME

ED 214 955 TM 820 106

AUTHOR	Quellmalz, Edys; Snidman, Nancy
TITLE	Evaluation Workshop I: An Orientation. New Edition.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
PUB DATE	[82]
NOTE	188p.; For related documents see TM 820 107-111.
EDRS PRICE	MF01/PC08 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; *Evaluative Thinking; *Evaluators; *Program Evaluation; *Training Methods; Workshops
IDENTIFIERS	*Center for the Study of Evaluation CA

ABSTRACT

The direction of this workshop is toward two major goals: (1) participant understanding of the kinds of information an evaluation can provide for educational decision making; and (2) participant understanding of the general procedures and problems involved in selecting, collecting, analyzing, and reporting that information. It is the intent of the workshop to provide a familiarity with essential evaluation skills. This is based on the objectives of the workshop; which are as follows: to distinguish the major characteristics of the Center For the Study of Evaluation (CSE) evaluation framework, and be able to properly categorize evaluation activities using the framework; to apply specific evaluation techniques which are appropriate for major evaluation activities in the CSE framework; to specify appropriate measurement devices, and design and analyses procedures for evaluation questions; and to interpret evaluation results and make informed recommendations to decision makers. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Center for the Study of Evaluation

UCLA Graduate School of Education
Los Angeles, California 90024

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

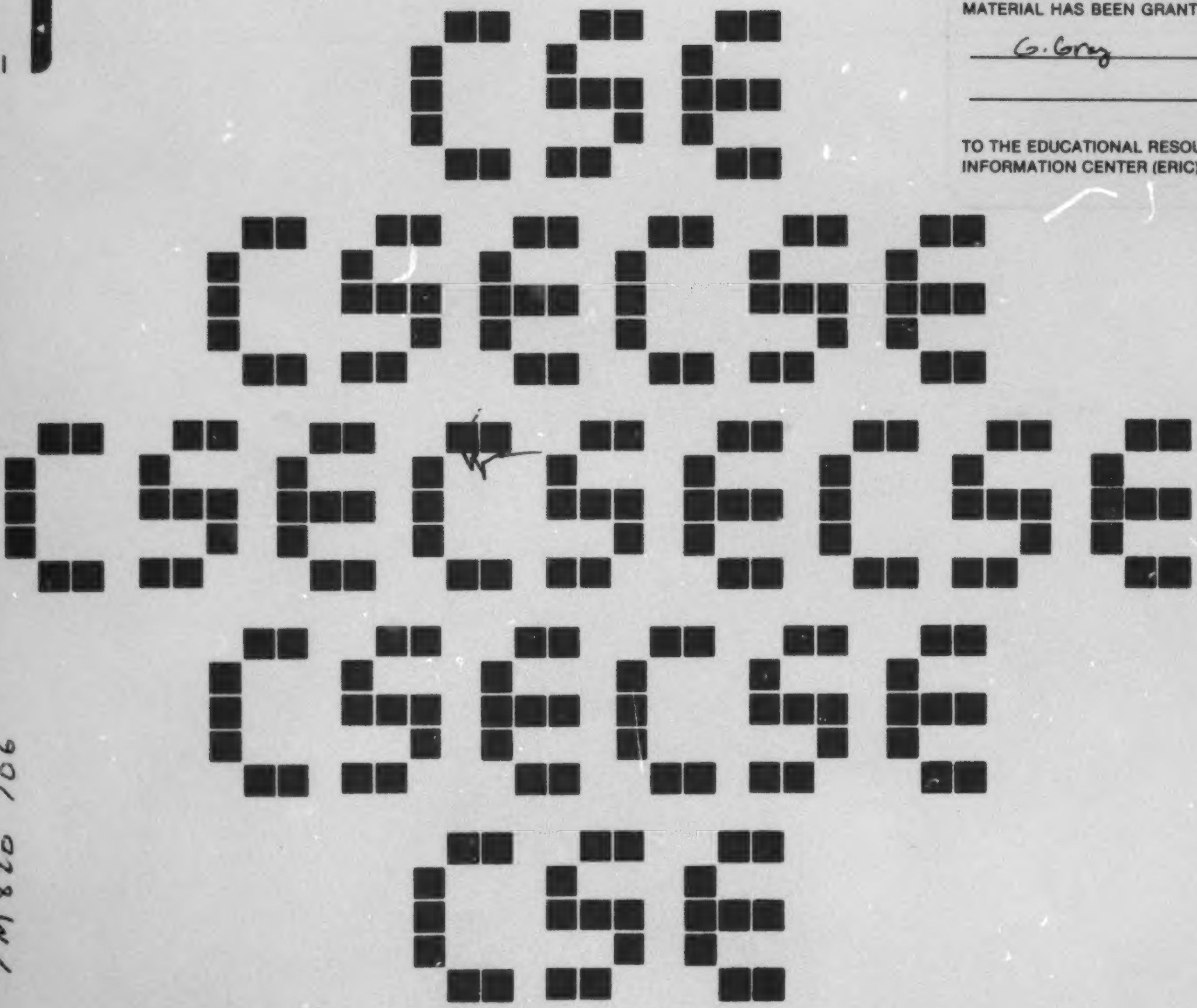
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED214955

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."



TM 820 106

Center for the Study of Evaluation

The mission of the Center for the Study of Evaluation is to conduct inquiry, from a variety of perspectives, into the nature of educational programs and services. Our commitment to inquiry into the field of evaluation grows from the belief that school practices and the competencies and satisfactions of those who participate in the educational enterprise can benefit from information collected in accordance with social science methodologies. Activities of CSE involve study of the instruments and methodologies for collecting information as well as the sociopolitical contexts of educational decisions as a means of contributing to the long-range growth in effectiveness of public education.

Information about CSE and its publications may be obtained by writing to:

Director, Public Information
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

24369

EVALUATION WORKSHOP I

AN ORIENTATION*

Edys Quellmalz
Project Director

Nancy Snidman
Project Coordinator

Developers: Marcella Pitts
Nancy Snidman
Kathi Yohn

Writers: Darrell Couturier
Rachel Grala

Editorial Assistant: Shelby Popham

---

*A new edition of EVALUATION WORKSHOP I: AN ORIENTATION by Stephen P. Klein with James Burry, David A. Churchman, and Marc Nadeau.

Printed in the United States

EVALUATION WORKSHOP I:

AN ORIENTATION

Contents

Section Page

Introduction to Evaluation Workshop I: Goals and Objectives i

A. CSE Evalution Framework A-1 - A-3

Exercise #1: Categorizing Evaluation Activities A-4 - A-5

B. Introduction to the Simulation

District Data B-1 - B-7
Simulation Characters B-8 - B-9
District Concerns Summary B10 - B13
Evaluation Context B14 - B20

C. Needs Assessment

Determining Program Goals and Objectives: Goal Section Procedures C-1 - C-5
Select the Goals Technique C-5 - C10
Exercise #2: Rating Objectives, Selecting Objectives C11 - C17
Needs Assessment Report C17 - C19

D. Program Planning D-1 - D-2

Context D-2 - D-4
Program Planners D-5 - D-7
Techniques D-7 - D17
Exercise #3: Decision Matrix Weighting Program Considerations D18 - D19
Programs Description D20 - D23
Section D Outline D24 - D26

E. Evaluation Planning

Purpose of the Evaluation: Formative and Summative E-1 - E-7
Context Considerations E-8 - E-9
Evaluation Questions E10 - E14

F. Measurement Instruments

Guidelines for Selection or Construction F-1 - F-3
Instruments F-4 - F11
Formative/Summative Considerations F11 - F12
Exercise #4: Selecting Appropriate Measures F13 - F14
Section Outline F15

Section Page

G. Evaluations Designs

Who G-1 - G-4
When G-4 - G-8
Formative/Summative Considerations G-8 - G10
Exercise #5: Choosing Designs G11 - G12
Section Outline G13 - G14
Planning the Evaluation G15

H. Data Collection

Guidelines for Data Collection H-1
Common Problems in Data Collections H-2 - H-3

I. Analysis Summarizing Evaluation Data

Descriptive Statistics I-1 - I-3
Tests of Differences I-3 - I-4
Tests of Relationships I-5 - I-6
Formative/Summative Considerations I-7 - I-8
Exercise #6: Data nalyses and Interpretation I-9 - I16

J. Reporting Evaluation Results

Outline of an Evaluation Report J-1 - J-6
Formative/Summative Considerations J-7 - J-8
Final Formative Report J-9 - J23
Exercise #7: Formative Report Summary J24 - J25

K. Feedback

INTRODUCTION TO EVALUATION WORKSHOP I

Goals

Evaluation Workshop I: An Orientation is directed toward 2 major goals.

- .Participant understanding of the kinds of information an evaluation can provide for educational decision making
- .Participant understanding of the general procedures and problems involved in selecting, collecting, analyzing, and reporting that information.

Objectives

Evaluation Workshop I is not designed to train people in all the skills needed by an evaluator. Instead, the workshop will give you a working familiarity with essential evaluation skills. The objectives of the workshop are:

- .Distinguish the major characteristics of the CSE evaluation framework, and be able to properly categorize evaluation activities using the framework
- .Apply specific evaluation techniques which are appropriate for major evaluation activities in the CSE framework
- .Specify appropriate measurement devices, and design and analyses procedures for evaluation questions
- .Interpret evaluation results and make informed recommendations to decision makers.

SECTION A

Section A

CSE EVALUATION FRAMEWORK

A. CSE Evaluation Framework
Exercise 1

This orientation workshop serves as an introduction to the four-stage evaluation framework developed by CSE. The four stages are: Needs Assessment, Program Planning, Formative Evaluation, and Summative Evaluation. One central idea underlies this evaluation framework: effective evaluation can aid educational decision making.

Thus, evaluation as discussed by these workshops, is a decision-oriented activity.

> Educational evaluation is the process of determining the kinds of decisions that have to be made; selecting, collecting, and analyzing information needed in making these decisions; and then reporting this information to appropriate decision makers. (Alkin, 1971)

This definition applies to each of the four stages of the Center's framework.

Table 1 summarizes each stage of the CSE framework by examining the:

.decision focus

.key activities

.reporting schedule

.role of the evaluator

.phase of program development

Table 1

CSE EVALUATION FRAMEWORK

Evaluation Stage	Decision Focus	Key Activities	Phase of Program Development	Role of the Evaluator	Reporting Schedule
NEEDS ASSESSMENT	Identify "high priority goals that will serve as the basis for new or revised educational programs	Select from among a full range of alternatives those goals to which resources should be committed. Determine extent to which current program is achieving the "high priority" goals	·goals/needs that should be addressed are to be identified ·a specific program has not yet been chosen	·help to clarify goals ·assist with identify-needs measure effectiveness of current programs vis-a-vis important goals	·when the needs assessment is completed
PROGRAM AND EVALUATION PLANNING	Decide the means for achieving previously selected goals	Select and/or de-develop the program most likely to attain goals. Develop a plan, including evaluation procedures, for the proposed program	·high priority goals have been identified ·a method for achieving goals must be selected ·evaluation procedures must be planned	·facilitate planning decisions ·help program planners develop, adapt or select program to meet high priority goals; build evaluation plan into program plan	·as the program is planned
FORMATIVE EVALUATION	Decide where improvements in an on-going educational program are needed	Measure program processes and components to determine feasibility of program plan. Assess achievement of program goals. Recommend improvments in program where needed	·a program has been selected and put into operation ·program is still "fluid;" modifications will be made	·determine extent to which program is running as planned ·measure program's progress in attaining its goals ·provide recommendations for improvement	·throughout course of the evaluation
SUMMATIVE EVALUATION	Determine the overall effectiveness of a fully-developed program to aid decisions about continuing, expanding, or dropping program	Measure program outcomes and document program processes to permit informed decisions about overall effectiveness of the program in relation to other programs	·program has been running for a while; procedures and methods are fairly stable (no major changes will be made)	·document program processes ·measure attainment of specified goals ·determine whether any unplanned effects have occurred ·compare new program to existing programs	·when the evaluation is complete

This table suggests an orderly and logical sequence for evaluation activities, beginning with Needs Assessment, proceeding through Program Planning and Formative Evaluation, and concluding with Summative Evaluation. This logic does, in theory, hold true. However, in actual practice an evaluator rarely has an opportunity to participate in all four evaluative stages for a single program. It is far more likely that the services of an evaluator (or several evaluators) will be called for at various points during the life of a program. Furthermore, the evaluator's mission will rarely be explicitly defined, such as: "We have now completed a needs assessment; will you please assist us in program planning?"

In order to circumvent the ambiguity inherent in many evaluation projects, evaluators must have a good understanding of the evaluation activities associated with different phases of educational program development.

Exercise 1

Categorizing Evaluation Activities

Directions

The Center for the Study of Evaluation has developed a four-stage framework model of educational evaluation. The four stages are:

1. Needs Assessment (NA)
2. Program and Evaluation Planning (PP)
3. Formative Evaluation (FE)
4. Summative Evaluation (SE)

During these stages, the individual(s) responsible for conducting an evaluation focuses on certain activities. In this exercise, you will be given 10 descriptions of evaluation situations. Your task is to select the evaluation stage into which each activity belongs.

Place the initials of the appropriate evaluation stage (NA, PP, FE or SE) in the blank before each activity.

EXERCISE 1

Categorizing Evaluation Activities

__________ 1. J. T. Kirk called in a communications expert Uhuru to determine the effectiveness of their well-established language program in Swahili in comparison to similar language programs.

__________ 2. The school board wants to discover what areas of the district's curriculum people feel are weak. The evaluator will distribute questionnaires to teachers, students, and parents to gather their opinions.

__________ 3. Since questionnaire data had indicated that his curriculum lacked the appropriate emphasis on canine tooth structure, B. Stoker called in Vlad Dracula, an evaluation specialist, to help district personnel create a suitable program to alleviate the problem.

__________ 4. The principal of Flinx High School, A. D. Foster, is primarily responsible for the school's new program on insect behavior. In order to determine whether or not this innovative program could use some modifications, he has summoned evaluator M. D. Pip to conduct an evaluation and offer frequent recommendations.

SECTION B

Section B

INTRODUCTION TO THE SIMULATION

Introduction to the Simulation
  District Data
  Simulation Characters
  District Concerns

Throughout the workshop, you will follow Lionel Edwards, an evaluator, as he conducts various evaluation activities in Hamilton Unified School District.

He will be working closely with Bruno Freitag. Freitag has been Superintendent of Hamilton Unified School District since unification eight years ago. Freitag's district contains two major communities, Victorville and Hamilton. This district is briefly described on the following pages.

DISTRICT DATA

•Description of Community. The district contains two major communities, Victorville and Hamilton, the latter being the county seat. The district is in the rural midwest and is made up mostly of people descended from families who came to the area during the mid-nineteenth century. Farming is the principal industry.

Two recent developments have had an impact on the kinds of students enrolled in the district's schools. The first development, a new junior college in Hamilton, brought in a comparatively large number of professional families who are quite vocal about how the schools should be run. The other development was the completion, two years ago, of a canning factory near the railhead in Victorville. This factory has increasingly attracted many low-income families from the state capital because of the relatively lower living costs, better working conditions, and higher wages in Victorville compared to those in the big city.

The district office and high school are located on the main road between Hamilton and Victorville. Each community has a junior high and three elementary schools. The district's Average Daily Attendance (ADA) is slightly over 11,000. The district has a single consolidated high school which serves 2,370 students, 900 sophomores, 700 junior, 770 seniors. This workshop focuses on the high school, especially its sophomore class. About 80% of the sophomore class can be expected to finish high school. Most dropouts leave school late in the tenth grade or early in the eleventh grade as they reach their sixteenth birthday.

.Parental concerns

To further acquaint you with the district, a conversation between two local parents, one from Hamilton and one from Victorville, appears below.

Characters: Chris (C) and Anna (A)

C Hi Anna, nice to see you.

A Hi Chris. Say, you and Paul haven't forgotten about coming over Saturday to see us in Hamilton, have you? Did you get a baby sitter?

C Yes, I did. But we almost didn't make it. Our usual sitter couldn't make it, so I called some young girl who advertised in the newspaper. But you know how many new families have been coming here to Victorville to work in the plant? Well, we don't know this girl or her family and she talked funny on the phone like she might be from one of the new families. I decided we don't want to leave Rosa with people we don't know. So Billy is going to stay home and watch his little sister. That was a real problem. Finally we had to just plain tell him he had to stay home.

A I'm glad you got it fixed up. I just don't understand those kids. It seems that we didn't have those problems a few years ago. But they sure are hard to handle.

C Paul thinks it's because of the new college over there in Hamilton. All those smart professors with their new ideas about education. Just teaches the kids a lot of nonsense. I mean they come here from the city and think they know what we need. For *our* kids. What do they know about living here? If father was still here, he'd run them off in a minute. Never did take to strangers who couldn't mind their own business.

A I know. Rudy met some of them at a town planning meeting and told them where to get off. But how about those people you're getting there in Victorville, the ones in the plant? We have some in Hamilton but it must be worse over where you are.

C It sure is. A lot of kids about Billy's age used to drop out of high school, but it's worse now. Those new kids are a bad influence. All they talk about is working and making money. Next thing our kids will forget all about having to run a farm some day and run off to the city. Why doesn't the school think about *that* for a change instead of all this new stuff?

A That's right. I remember when we had to take things like science and biology they made it sound like it would help you know the farm and things, but not any more. All the kids want to take up and leave, it seems.

C Yeah, I know. Billy doesn't even get as good grades as he used to. He's staying out late at night with some of those new kids in the high school. I wish they'd close up that plant and they'd all go away.

A I wish they'd take those new college teachers with them--and their beards and their funny clothes. Most of them look like they couldn't lift a sack of potatoes unless they got some help from a small boy.

C And their wives are even worse. What about the language they talk? And our kids have to hear that. No wonder they're different now.

A I know. And from our money. I can't even buy a new dress anymore it's so bad. Listen, I have to go now. So we'll see you about seven o'clock. OK?

C That's fine, Anna. I'm looking forward to it. See you on Saturday. Bye.

A Bye.

.Excerpts from Local Newspaper

Page 4 Hamilton Enquirer October 12, 19XX

TELLING WEEDS FROM CORN

We have been studying our local schools for some years now. It doesn't take much to see that they are costing more every year. Let's look at some facts. Every year we put 30% of our kids into college, 10% into the army, and 15% into downstate trade schools. Every year 20% of the sophomore class drops out.

If the schools were getting better you would expect these numbers to be changing. Especially the number who go to college and the number who drop out. But they haven't changed for the ten years that we have been watching the schools from this newspaper office.

And what are we getting for our money? It costs an average of $400 per year for 13 years for every kid we graduate. That comes out to over $3,500,000. What do we get in return for that three and a half million dollars?

Thirty percent of each graduating class goes to college. Most of them finish. But only about 15 of them come back to make their living here. Each year about 70 of our young men join the army. How many of them do we see again? As the old expression has it, "How are you going to keep them down on the farm, now that they've seen Paree?"

What it all comes to is that of almost a thousand kids who started school 13 years ago, about the only ones who will remain in the area are the 200 or so who dropped out before graduating, and the ones that don't have much in the way of courage or initiative or imagination. These are the kids who will inherit the land that was pioneered by our forebears. The schools *must* develop programs that will instill in our brightest children that love of the land which made their fathers the backbone of this country.

The schools keep telling us that they have to make schooling *relevant* if the kids are going to learn, and we have to pay a fortune for new-fangled do-nothings to make school relevant. And yet we have kids raised in the finest farmland in the world who either want to run off to live in the cities, or if they stay here, they can't tell a bull from a cow, tell weeds from new corn, or rig a pulley to get hay into a loft. *These* are the things that are relevant here; they should be taught in a way to make them relevant to the kids. That's the only way we will be able to keep the graduates of our schools here to raise their families and strengthen the local economy.

LETTERS TO THE EDITOR

Dear Editor:

Impressed as I am with the possibilities for change here at the college, I am dismayed at the demands of so many of the "local yokels" who seem to want to keep themselves and their children in the agrarian dark ages. What we really need is a higher quality education for our college-bound youth who must be able to meet the demands of modern society.

Dr. J.R.
Hamilton

Dear Editor:

Keep up the good work! Like the vice-president says, what we need is more vocational training and less college for kids. That's true in high schools too. Kids don't appreciate anything anymore.

E.K.
Victorville

Simulation Characters

The cast of characters and the role they will play in the evaluation are described in the personnel profiles which follow:

Bruno Freitag

Age: 45

Position: Superintendent, Hamilton Unified School District

Education: M.A., Administration, State University
B.A., Education, State University

Experience: Superintendent, Hamilton Unified, 8 years
Assistant Superintendent, Hamilton Unified, 4 years
Principal, Malden Jr. High School, 5 years

Comments: For the last few years Freitag has focused on local problems, particularly high school dropouts and the rising cost of education in the community. He is not overly opposed to new ideas in education and has attended a few workshops which exposed him to some of these innovations. He sees values in some of the new programs but is still fairly "traditional" and worries about the effects of change when it is introduced too quickly.

Freitag's ancestors were among the first families to settle in the district. His family has been active in farming for several generations.

Sandi Silverthorn

Age: 32

Position: Curriculum Specialist

Education: M.A. in Curriculum, State University
B.A. in Education, State University

Experience: Curriculum Specialist, 3 years
Curriculum Consultant for Grant Unified School District, 5 years
Teacher, Togliatti Junior High, 2 years

Comments: In her role as Curriculum Specialist, she has participated in many seminars, graduate courses, and workshops in educational innovation. She is much more responsive to change than Freitag.

Silverthorn is single. The fact that she has never had children often creates a problem for her when she talks to parents about new ideas in education. The community considers her ideas *avant garde*. Silverthorn's family came to the area early in the century and soon established a successful business in farming supplies.

Evert Walker

Age: 52

Position: Principal, Hamilton High School

Education: B.S. in Agriculture, State University

Experience: Principal, Hamilton High School, 12 years
Principal, Wagner Junior High, 10 years
Teacher, Wagner Junior High, 7 years
Substitute Teacher, 3 years

Comments: Since Walker has been the principal, test scores at Hamilton High have always averaged in the top half of those reported for the state. He is a very assertive principal and demands that all teachers channel their administrative functions, including trivial requests for supplies, through him. He is suspicious of people he does not know and resents outside interference in his school. Walker also comes from a family who has been in Hamilton for many generations.

Lionel Edwards

Age: 31

Position: Evaluator, State Evaluation Agency

Education: Ph.D., Educational Evaluation, UCLA
M.A., Education, Columbia University
B.A., English, Columbia University

Experience: Intern with American Association of School Administrators - 2 years
Teacher - 10th grade English
Delaware - 1 year

Comments: Dr. Edwards has not only a theoretical background but also a practical knowledge of the application of evaluation theory and methodology. During the last two years, he worked as an intern with the American Association of School Administrators where he evaluated a reading program for inner-city schools while compiling data for his doctoral dissertation.

As an evaluator with the State Evaluation Agency Edwards conducts evaluations in the southern region of the state. Edwards is married and has 2 sons, both attend elementary school. His wife teaches third grade.

.District Summary

The editorial you read earlier and the conversation you heard voiced many of the important community concerns about Hamilton Unified's educational program.

To review, some of these concerns were about dropouts, the number of students attending college and their preparation for it, the seeming high cost of education relative to what the students are getting for it, and the apparent inappropriateness of the curriculum relative to the needs of the community.

These concerns, plus the legislative pressures to make education more accountable for its funding, plus the impending battle between Hamilton and Victorville parents regarding what should be taught in the schools, led superintendent Freitag to the conclusion that he had better get some accurate information about what was needed in his schools before going ahead and planning new programs.

At the suggestion of another superintendent, Freitag called the state department's educational evaluation agency to see whether they could help him in gathering the information he needed to make a decision about new programs. Dr. Susan Lightner of the agency worked with Freitag and helped him *select* the kind of information he needed and showed him how to *collect* and *analyze* that information. For example, she explained how to develop appropriate statements of possible goals for the district and how to sample various members of the community, teachers, and students in order to obtain their views about what goals are the most important. Lightner also helped in determining the extent to which some of these important goals were being met so that Freitag could identify which areas really required new programs. In short, Lightner helped conduct a needs assessment for the district.

The needs assessment was completed last spring. Now Freitag and the school board will use the needs assessment results to help make decisions about which school programs should be changed. Attention will be focused on ways to improve existing programs or develop new ones that will help fulfill identified needs.

Even though the needs assessment indicated several areas that needed improvement, there is not enough time in this workshop to follow the district as they work toward achieving all their goals. Instead, the workshop will limit its scope to a single goal. Since one result of the needs assessment was the allocation of money by the school board to correct the deficiency in tenth grade biology at Hamilton High, the attainment of this goal will be the focus of this workshop.

The superintendent, Bruno Frietag, has decided, with agreement from the school board, to hire an evaluator to assist in all phases of the biology program improvement. Although Susan Lightner did the initial needs assessment, she will be unable to work with the district on the subsequent evaluation activities since she was promoted to Director of Field Service Unit at the Evaluation Agency. Instead, Lionel Edwards, who recently joined the agency, will be working with the district.

Frietag has contacted Edwards and an excerpt from their first meeting is found on the following pages.

Characters: Freitag (F) and Edwards (E)

F: Dr. Edwards, how nice to meet you.

E: Mr. Freitag. Please, call me Lionel.

F: Yes, and I'm Bruno. I've heard a lot of good things about you. I'm looking forward to working with you.

E: Well, thank you. Susan Lightner had quite a few complimentary words about you too.

F: Ah yes, Susan. Haven't heard much from her since she became the director of the field services unit of the State Evaluation Agency. However, I did speak to her briefly a little while ago about getting someone for this job. She strongly recommended you for the assignment. Well, enough about Susan. Let me explain exactly why we need you. We are experiencing a lot of pressure from various factions about our educational program. As you know, Susan conducted a general needs assessment of our district before she assumed her current position.

E: Yes, what did the assessment turn up?

F: Essentially five areas in which we have the greatest need for improvement: reading in the primary grades, mathematics in the upper elementary and junior high grade levels in Victorville, and, at Hamilton High, tenth-grade biology, eleventh-grade English, and twelfth-grade history. A number of other high priority goals deserving attention, such as reducing the number of dropouts and increasing the number of graduates going to college, were also identified in the needs assessment.

E: I see. Earlier, you referred to pressure that you are receiving. . . . What kind of pressure is it and who's exerting it?

F: Oh, the parents are concerned about dropouts, the number of graduates who attend college and their preparation for it, the high cost of education as it relates to the quality and amount of instruction received, the discrepancy between what the kids learn in school and what the community feels they should learn for "real" life, and so on. Those are a few of the concerns.

E: Right -- uh, you say those are the parents' concerns. Any other sources?

F: Of course, we must not forget the recent legislative pressures to make education accountable for its results and funding. So . . ., I figure, between the parents and the legislature, I'd better get some accurate information about what we need in our schools before planning any new programs.

E: Under the circumstances, it does seem the wisest course.

F: So, now that we've identified our basic needs, I and members of the school board have decided one area in which to concentrate our efforts is the biology program deficiency at Hamilton High. We have some money available that could be used to purchase a new program. First we would like you to help us conduct a needs assessment of the current biology program.

E: Okay, fine. I'll tell you what, why don't we . . . . . .

(FADE OUT)

EVALUATION CONTEXT

Before Lional Edwards begins a needs assessment to select curricular goals for the biology program he will have to become familiar with the context in which he will conduct the needs assessment. The next few pages highlight some of the important questions Edwards should ask to get a sense of the context in which he will be working.

It is very important that the evaluator be aware of and understand the context of the program to be evaluated. Context refers to all background and situational information that may have an effect on the program and the evaluation of the program. Contextual information includes social and political concerns as well as methodological ones. These include:

.school organization and structure
.community organizational structure
.communication flow
.role of the evaluator
.program characteristics
.evaluation requirements

To clarify the evaluation context, the evaluator asks questions pertaining to each of these areas. The specific questions asked depends on the stage at which the evaluator enters (needs assessment, program planning, formative or summative evaluation). Since Edwards is entering at the very beginning of the evaluation process, before a needs assessment has been conducted, information pertaining to some areas does not yet concern him (e.g. program characteristics).

Four areas which do concern Edwards are:

- .school organizational structure
- .community organizational structure
- .communication flow
- .role of the evaluator

Listed below are some questions pertaining to each of these areas. These questions are not all inclusive but are meant to serve as examples of relevant questions the evaluator might ask:

.School Organizational Structure

A. Who are the decision makers?

1. What is their role in the organization?
2. Why are they involved in the needs assessment evaluation? (self-initiated or required?)

   What is their knowledge of evaluation?

   What is their commitment or attitude toward evaluation?

   What is their managerial style? (shared, authoritarian, timid, aggressive).

   What are their backgrounds? (education, work experience)

B. Who are the school administrators?

1. How strong is their support?
2. How strong is their evaluation expertise?

.Community Organizational Structure

A. What involvement with the school do parents, businessmen and other community persons have?

1. What attitudes toward the school do they have?

    2. What information and attitudes toward evaluation do they have?
    3. What will be their involvement in the needs assessment evaluation?
B. What skills (education, work experience) do those from the community who will be involved in the needs assessment possess?
C. What political conditions in the community, district or state exist that may influence the needs assessment evaluation activities? (e.g., is state funding pressure creating a need for change in the school district's goals?)

.Communication Flow

A. How frequently should reports be made?
B. What audiences are there for interim needs assessment reports?
C. What are the audiences for the final needs assessment evaluation report?

.Role of the Evaluator

A. Is the evaluator internal or external?
    1. part of the school/district staff
    2. a third party outsider
B. Is the evaluator to be helpful, nurturing, or neutral and objective?

.Acquiring Contextual Information

Edwards has received answers to some of these contextual questions in his initial conversation with Freitag. In order to uncover more background information, he has decided to contact his director, Susan Lightner, who conducted the general needs assessment in Frietag's district before her promotion to her new position. She should be able to provide him with some of the inside information on the district and people involved in the needs assessment.

The following is a transcript of Lionel's telephone conversation with Susan:

Characters: Lionel Edwards(E) and Susan Lightner (L).

(Telephone rings)

L: Hello

E: Susan? This is Lionel Edwards

L: Lionel, hi!, how are you?

E: Fine. I'm just getting into something that I could use your help on. You know that evaluation job that you recommended me for in Freitag's district?

L: Yes.

E: Well, considering your past experience as an evaluator in that district, I thought I'd call you and ask you a few questions about the district.

L: OK, fine. What did you have in mind?

E: Well, your general needs assessment turned up a deficiency in the 10th grade biology program at Hamilton and they've decided to concentrate on that problem. So, I'm to conduct a needs assessment of the biology program. I've had a meeting with Freitag at which he laid out the general tenor of the district -- pressures, attitudes, etc. -- and what he wants me to do. Seems like an amiable type.

L: You're right. He's quite nice and cooperative. Although he makes a concerted effort to be open to innovative programs, I sometimes feel that underneath it all he's a bit of a traditionalist. Maybe his biggest worry is just his concern about upsetting his district with changes that are too radical and too sudden. If you just don't pressure him, you'll get great results. And get your reports in on time, he's very traditionalist in that respect.

E: OK. How about the community? When you worked with the parents and business people did you find them supportive, interested or knowledgeable about the schools in general?

L: As Freitag must have told you, there are several diverse groups in the two communities including factory workers, university people, and farmers. It would probably be a good idea for you to set up a committee for the needs assessment phase that includes representatives from the various communities, business sectors, and schools. Perhaps that way you can lessen any friction between your work and new programs and the parents, business people, and teachers. Bear in mind that your needs assessment report will be read by community people. Protect yourself and your findings from any possible antagonism by keeping accurate records and logs of everything. That way you will be able to substantiate your findings should they be challenged.

E: I see. Well, listen, Susan, you've been a big help. I'll call you again and let you know how I'm doing.

L: Fine. Let me know if you need any help and please send me a copy of your letter of agreement.

E: OK. Thanks a lot. Talk to you later.

L: You're quite welcome. Bye.

E: Goodbye.

Following this conversation Lionel wrote the following memo to Frietag in which he briefly outlined his intentions for the needs assessment.

March xx, 19xx

TO: Bruno Freitag

FROM: Lionel Edwards

RE: Needs Assessment (Biology Program/Hamilton High)

Dear Mr. Freitag:

I would like to take this opportunity to formally accept the evaluation position that you offered to me.

As I understand it, the needs assessment will last _____ months. The final report to be submitted by ______________________.

The needs assessment I would like to set up includes a committee composed of two parents from each community, two teachers and the principal from each school, and two representatives from the business sectors. If you wish to add anyone to this committee please contact me or we can discuss it at our meeting on Wednesday of next week. At that time I will submit a tentative evaluation budget for your review.

I am confident that we can finalize plans and draw up a formal contract within the next two weeks.

Sincerely,

Lionel Edwards

SECTION C

Section C

NEEDS ASSESSMENT

C. NEEDS ASSESSMENT
- .Determining Program Goals and Objectives: Goal Selection Procedures
- .Select the Goals Technique
- .Exercise 2
- .Needs Assessment Report

Lionel Edwards has signed a contract with the Hamilton Unifed School district to conduct a needs assessment for the 10th grade biology program. Needs assessment is the first stage in the life cycle of a program; the initial step in the CSE evaluation model. The impetus for needs assessment usually arises froma necessity to determine priorities in order to allocate human and material resources.

The purposes of a needs assessment are:

- .to determine what goals and objectives are desired
- .to obtain data on how well these goals and objectives are being met by the existing programs
- .to prioritize goals and objectives for which a descrepency exists between desired and actual outcomes

Determining Program Goals and Objectives: Goal Selection Procedures

One of the major purposes of a needs assessment is to determine the goals a program should try to achieve. Various goal selection procedures include:

VALUES CLARIFICATION TECHNIQUE

The values clarification technique takes a principally affective and introspective orientation to generating educational goals. Values clarification is most useful for stimulating thought about general goals of schooling rather than for developing specific objectives. Other, more traditional, goal-preferencing methods focus on concrete problems or pre-prepared goals--mainly cognitive goals. Using a single meeting format, the values clarification technique invites adults to speculate in a series of exercises about what they value most highly as qualities in themselves and in their children. A needs assessment based on this technique gives the school direction about the goals it should pursue. Using this method, participants generate their own goal statements. These will generally turn out to be quite broad; they may be conventional or innovative, easy or difficult to implement. In any case, the values clarification process will stimulate discussion about the long-term purposes of schooling.

The values clarification technique asks for more thought and self-examination than do some other goal-preferencing methods. It requires that teams of participants talk about themselves and finally describe and prioritize qualities they value in educated people.

PROBLEM SNAPSHOT TECHNIQUE

The problem snapshot technique uses participants' knowledge about current and past problems as a basis for suggesting school goals. The focus on problems is what is unique about this needs assessment technique.

The problem snapshot approach presumes that when people describe school related problems, they perceive a situation which is at variance with what they desire. It is assumed that individuals have done their own private and personal needs assessments, and what remains to be done is to combine those private and personal perceptions into a public consensus about the issues which should be addressed by planners.

The problem snapshot technique for goal preferencing accomplishes a needs assessment through two separate sequential activities:

1. Participants in their homes or offices examine a pre-provided Problem Checklist and write down what they consider to be salient difficulties hindering their school's or district's accomplishment of its objectives. Participants are free, of course, to add problems not listed. The problems are returned to the evaluator. The evaluator summarizes the responses received and prepares a list of most frequently identified problems to be discussed at a needs assessment meeting.
2. The needs assessment meeting provides a forum for presentation of the problems and conversation about them. Its goal is to select from the preidentified problems those which the school or district should tackle. Participants as a group are asked to contribute problems not listed, then small teams convene to prioritize the problems according to their urgency and to the possibility of action to solve them.

FUTURING TECHNIQUE

The futuring technique was devised in order that those concerned with planning can take advantage of a method of generating long-range educational goals. Other goal-preferencing methods focus on current problems, present values or currently-explicated goals. The futuring technique uses imagination and speculation to project a view of the world the students will inhabit as adults. From this perspective, participants work backward to create present-day goals. These goals may be global or broad, traditional or original, easy or difficult to implement. In any case, the futuring process will stimulate discussion about the long-term outcomes of schooling.

During a single meeting, participants list out areas of life concerns and hypothetically explore the status of these areas 15 years hence.

POSTAL POLLING TECHNIQUE

Postal Polling is a process for organizing and sharing ideas, and ultimately arriving at consensus through a series of mailed questionnaires. The purpose is to gather together the thoughts of people who cannot attend a needs assessment meeting because of time or distance difficulties.

This approach is most suitable for situations where time is not severely limited, since response time will increase with the number of participants.

Each participant responds to an initial mailed open-ended questionnaire. Participants then receive feedback about how others responded and are given the opportunity to revise or expand upon their original reactions. A second tabulation of responses is then submitted to all respondents. From this procedure emerges consensus concerning the answers to the original questions or a clarification of the differences in position among the respondents.

RATING OBJECTIVES TECHNIQUE

Since the Rating Objective technique is one of the most frequently used needs assessment strategies, and the one Edwards has decided to use in his needs assessment of the biology requirements at Hamilton High, this technique will be discussed in some detail.

In this strategy the evaluator guides three main activities:

- .assembling an initial pool of goals and objectives
- .selecting raters
- .developing and implementing a process to rate objectives

•Selecting Objectives

Objectives can be either selected from a number of prepared sources of objectives or generated from a number of different information sources. Some of these are listed in Table 1.

Table 1

Potential Sources of Objectives

- .State and federal mandates and legislation
- .Local and national concerns identified in editorials, articles, and legislation
- .Community concerns voiced by business organizations, pressure groups, etc.
- .Parental concerns as voiced in letters, PTA meetings, and conferences with parents
- .School records and reports, such as attendance, discipline, career choices of graduates, and test scores
- .Curriculum guides, program descriptions, and recommendations of responsible educational institutions and organizations, such as the National Science Foundation, National Assessment Project, NEA, etc.
- .College requirements and requirements for graduates, such as employers' recommendations
- .Teachers' reports and comments
- .Student requests and comments
- .Prepared composite sets of objectives and goals, such as may be obtained from the Instructional Objectives Exchange and the Elementary School Evaluation Kit.

A sufficient number of goals and objectives should be compiled to ensure adequate coverage of the subject matters.

When possible, goals should be accompanied by objectives stated in behavioral terms so that the rater has a clearer idea of the behaviors toward which an objective is directed. While a needs assessment could begin with various groups rating general goals, another round of ratings would be necessary once objectives clarifying the intent of the goals were sepcified.

A list of objectives should be prepared. One way of doing this is to print the objectives on cards, one objective per card. It is helpful to describe an objective according to: 1) the subject area and 2) the general goal. Each objective can be numbered for ease in recording the ratings.

The source(s) of the objective may be printed on the card; e.g., State Curriculum Guide, National Science Foundation, recommended health program, etc. These sources may be especially important when there are legal constraints, state directives, or university requirements that must be considered. An illustration of a goal and sub objectives for a biology program appear below:

#12 Subject: Science.........

Goal: To understand man's relationship to animal and plant organisms in his environment.

Objectives derived from goal

A. Can describe changes in the environment implied by changed plant or animal conditions

B. Can describe functions of various plants and animals.

C.

D.

Source: State
Curriculum Guide

·Selecting Raters

Edwards has compiled a list of biology objectives. After consulting with Frietag, Edwards is ready to select the needs assessment committee. This is a critical step since the viewpoints of the raters will determine which objectives will finally be selected. It is generally best to use various kinds of raters, such as teachers, parents, students, school board members, curriculum experts and consultants, community pressure groups, principals, district administrators, etc. The type and number of raters from each category is a function of the size and focus of the task. For example, it would be more important to involve the community in selecting general goals for curricular areas than in rating the objectives for a tenth grade biology course. On the other hand, the advice of biology curriculum experts will be more valuable in rating biology objectives than in rating goals in other curricular areas. Edwards, and Freitag will consider these factors in selecting raters.

·Sampling

Although it is better to include too many people than to omit someone who should be involved, it is often not practical to include everyone in the ratings, such as the parents of all children in the school district. If Edwards wishes to limit the number of people involved, it is possible to sample randomly from the total group of potential raters in a systematic manner. A detailed discussion of sampling procedures is contained in the design section of the workshop.

·Weighting

In some instances, it may be desirable to weight the ratings of each kind of rater. This involves deciding how much emphasis to place on the views of each group of raters. The simplest system for applying weights is to use whole numbers. For example, if it was decided to give parents twice the weight given to taxpayers without children, then the weight for parents would be 2 and the weight for taxpayers would be 1.

The selection of raters and the weighting procedure can become an important political issue. It is wise to have many different groups represented when these decisions are made.

·Rating Objectives

Once the objectives and raters have been selected, the evaluator is ready to have the objectives rated. Everyone should rate the objectives on the same scale of importance. A five point scale like the one shown below might be used

1 = Unimportant, inappropriate, or irrelevant
2 = Below average importance
3 = Average importance
4 = Above average importance
5 = Very important, critical, or essential

The evaluator compiles the ratings and computes an average rating for each kind of rater. For example, if there were only 3 teachers and their ratings of the first objective were 3, 4, and 5, then the average (or "mean") rating for the teachers on the first objective would be 4.0 (3+4+5=12; and

12 ÷ 3 raters = 4). These figures are shown in Table 2 along with parents' and students' mean ratings for the first and second objectives.

Table 2

Sample Roster of Ratings from Three Groups

Objective Number	Teachers (N=3)					Parents (N=4)						Students (N=2)			
	1	2	3	Total	Mean	1	2	3	4	Total	Mean	1	2	Total	Mean
1	3	4	5	12	4.0	4	5	5	5	19	4.8	3	4	7	3.5
2	1	1	2	4	1.3	4	4	5	5	18	4.5	2	3	5	2.5

Exercise 2
Rating Objectives

In this exercise you will have a chance to practice the Rating Objectives technique. Twelve of the objectives for tenth grade biology are provided on the following page. These are just a sample from the larger pool of objectives rated by the committee.

Each team will assume the role of one of the three interest groups in the Hamilton Unified School District--farmers, canning factory workers, or university persons. All members of a team will be members of the same group.

Directions:

1. Working individually, assume your role, and rate each objective by assigning it to one of the following 5 classes of importance:

5= Very important, critical, or essential

4= Above average importance

3= Average importance

2= Below average importance

1= Unimportant, inappropriate, or irrelevant

Record each rating in the first column following the objective statements marked "Rating".

To assure that the objectives are spread out, put at least two objectives in each numerical rating category.

2. Compare the ratings with those of your teammates. Compute the average rating (sum the individual ratings and divide by the number of people in the team). Record the team rating in the column marked "Team Rating". Be ready to report your team's top three objectives.

RATING CHART

Goals/Objectives	Rating (1-5)	Team Rating
1. Students will explain various interactions between organisms and their environment		
2. Students will develop the ability to deliver minor first aid		
3. Students will explain or demonstrate various chemical processes necessary for life		
4. Student will demonstrate improved personal health care and hygiene		
5. Students will display a more positive attitude toward and greater interest in biology		
6. Students will describe the structure of an organism and state how its structure relates to its function		
7. Students will display improved proficiency in various laboratory techniques and in the use of laboratory equipment		
8. Students will list strategies to minimize environmental pollution		
9. Students will classify organisms according to phyla and list the characteristics of various phyla		
10. Students will describe the growth cycle of plants		
11. Students will correctly describe specific processes in reproduction and tell how they contribute to the continuity of life		
12. Students will demonstrate an understanding of the role and importance of food and nutrition		

Collecting Performance Data

Once a goal-preferencing technique has been used to select goals, a second major activity in needs assessment is obtaining data on how well the selected goals and objectives are being met by the existing program. Now that the group of raters has completed the selection of important goals for the biology program at Hamilton High School, Lionel Edwards' task is to collect performance data on the present level of achievement relative to the highly rated objectives.

Since it is unlikely that information could or should be collected on all important objectives, those objectives on which data should be collected must be determined. In determining these objectives, various factors must be considered:

- Rated importance of the objective. The higher the final average rating, the more likely performance on it should be assessed.
- Cost of collecting and analyzing the data. It may be better to assess performance on five moderately important objectives than on just one very important one if limited funds require making such choices. The data may allow decisions affecting more aspects of the curriculum.
- Quantity and quality of data to be collected. Is the available sample large enough to ensure reliable data? Can the data be complete for all individuals in the sample? Will the data collected be credible or will its quality cause it to be considered questionable?
- Feasibility of collecting the data. The time and resources needed to collect the information must be considered. For example, is a computer necessary? Is it legal to use a particular measure? Are

the necessary trained personnel available to do it? Does it take too much student time? Etc.

Desired Performance Level

Information about student performance on highly rated objectives is collected to determine the size of the discrepency between actual and desired performance. Each objective, therefore, should contain a statement of "desired" performance levels.

The process of setting desired performance levels, however, is a difficult one at best. It is often made even more difficult by having to set standards for various subgroups, such as students in different tracks and grade levels.

Setting performance standards is especially difficult in the area of affective objectives, although clear statements of desired behavior can help to alleviate the problem.

Standards on each objective can often be set more easily after determining how performance on them will be measured. Even before a needs assessment is formally begun, present performance levels may be known through such means as the publication of state test scores. Test scores often stimulate a needs assessment in the first place. When interpreted properly, such information may be used in providing a frame of reference for setting realistic performance levels.

Student performance data may be collected from a number of sources. Edwards will use a variety of measures and data collection methods. Measures and data collection are discussed at length in sections F of the workshop.

Final Selection of Objectives

After collecting performance data, Edwards needs to establish procedures to decide which of the discrepancies between present performance and desired performance on highly rated objectives are most important to correct. This step requires two types of information:

- The rated importance of the objective. (Ratings were obtained from the data selection stage of the needs assessment.)
- The discrepancy between present and desired performance levels on each objective.

Each discrepancy between desired and actual performance is rated as being "Small," "Medium," or "Large."

This requires a series of somewhat arbitrary decision rules. For example, one decision rule might be as follows: When the desired performance (such as ratings of sportsmanship) occurs over 75% of the time, the discrepancy is small; when it occurs between 50% and 75% of the time, the discrepancy is medium; and when it occurs less than 50% of the time, it is large. Such rules should be constructed for all measures prior to the data collection. Next, the discrepancies, large, medium or small, must be quantified. One way to quantify discrepancies is to make use of a decision matrix like the one pictured on the next page.

DECISION MATRIX

Importance of Objective	Size of Discrepancy			
	Large	Medium	Small	None
5.0 4.0	6	5	4	3
3.9 3.0	5	4	3	2
2.9 2.0	4	3	2	1
1.9 1.0	3	2	1	0

In this matrix, the rating of an objective is cross tabulated with its discrepancy size to obtain a score. To use the matrix, first look in the lefthand column "Importance of Objective" for the final rating given by the committee to an objective. Then look across to the box giving the appropriate size of discrepancy. The number in the box is the final score. For example, an objective with a rated importance of 3 and a medium discrepancy should receive a priority score of 4. The numbers in the box are arbitrary and only serve to allow a spread of numbers to simplify final selection of objectives.

With a final score for each objective an easy comparison can be made of the relative priorities of all objectives. The program planners would use this information to begin planning programs that will help achieve the high priority objectives. In developing such plans, however, planners must not lose sight of the necessity to maintain present performance on those objectives that were rated high in importance and were being adequately achieved (i.e. did not show discrepancy between desired and actual performance.) In addition they should consider practical problems dealing with the feasibility and cost of implementing various objectives as they finalize their list of program objectives. Many of these considerations are reflected in Lionel Edwards' Need Assessment Report which follows.

Needs Assessment Report

To conclude the need assessment phase of the evaluation, Edwards prepared and submitted the following report to Bruno Freitag.

NEEDS ASSESSMENT REPORT
Biology Program at Hamilton High School

1. Preparation of a comprehensive set of objectives.

Thirty-five objectives for a tenth-grade biology program were selected from a comprehensive set of objectives for high school biology prepared by the state. The committee selecting the objectives was composed of Lionel Edwards, the six biology teachers and six parents. The objectives were printed on 3x5 cards for easy sorting into categories using the Rating Objectives technique.

11. Determining the relative importance of the objectives.

A one-day workshop was held at Hamilton High School during which one hundred participants rated the objectives. Raters from Victorville and Hamilton were equally represented and included thirty-four teachers from Hamilton High School, thirty-four parents, fourteen students, ten community leaders and eight administrators or curriculum specialists.

The raters had equal weight in the rating procedure. A final composite rating was computed for each objective. The 8 highest ranking ed objectives were:

1. Students will display a positive attitude toward and interest in biology.
2. Students will explain various interactions between organisms and their environment.
3. Students will explain or demonstrate various chemical processes necessary for life.
4. Students will classify organisms according to phyla and list the characteristics of various phyla.
5. Students will describe the structure of an organism and state how its structure relates to its function.
6. Students will list strategies to minimize environmental pollution.
7. Students will correctly describe specific processes in reproduction and tell how they contribute to the continuity of life.
8. Students will develop the ability to deliver first aid.

III. Identifying discrepancies

Performance standards were set by consensus of a representative committee from the initial rater group consisting of two teachers, two parents, one student, one community leader, and one administrator. State and national test norms were used where applicable. It was found that 80% of the students were eligible for certification by the Red Cross in first aid skills and that 85% of students completing tenth grade biology could list strategies to minimize environmental pollution. Student performance on the remaining objectives was below expected standards.

IV. Final selection of objectives.

The remaining 6 objectives were rated as having medium or large discrepancies between performance and standard. The committee decided to plan a program to help students achieve these objectives. The program will include instruction to maintain achievement on objective number 6. First aid instruction has been and will continue to be provided by various community groups including the YMCA, Boy and Girl Scouts, and the American Legion. Action on the other lower ranking objectives will need to be deferred until more time and money are available.

SECTION D

SECTION D

Program Planning

Program Planning
  Context
  Program Designation
    Tools
  Exercise

Program Planning involves developing, selecting or adapting a program.

Now that important biology objectives have been identified, a program to attain these objectives can be planned.

The evaluator helps in the development, selection or adaptation of a program and takes major responsibility for planning its implementation and evaluation.

The evaluator's specific tasks vary depending on whether a program is developed or selected.

.If the program planners have decided to develop a program, an evaluator would:

.help planners clarify the objectives

.examine relationship between objectives and planned activities

.examine relationship between measures and the objectives and planned activities

.allocate time, money and resources for program development

If program planners decide to select a program from among existing programs, the evaluator will perform different tasks. Since Hamilton Unified has neither the time nor resources to develop its own program, Edwards will help the district select a biology program whose objectives closely match those of the district. Edwards' specific tasks include the following:

.analyze and summarize existing programs so that information about them can be compared

.help planners list program characteristics they feel are important

.suggest techniques to choose programs

.evaluate the information reported by each program pertaining to development and evaluation

.evaluate the adequacy of the testing, diagnostic, and record keeping instruments included in the program

THE PROGRAM PLANNING CONTEXT

If an evaluator is just becoming involved in an evaluation during program planning and has not "been on board" during needs assessment, it is important to gather background data about the project. In the case of this simulation, however, Lionel Edwards conducted the biology needs assessment for the Hamilton Unified School District, so he already has information concerning the school and community. Before beginning the program planning stage, he must also determine specific program planning requirements:

.What is the purpose of program planning?
    .program selection?
    .program development?
    .program adaptation?
.What resources are available for program planning?
    .time and money allotted for planning?
    .personnel who will participate in program planning?

At every stage of the evaluation it is wise for the evaluator to summarize and document this additional contextual information and present it to decision makers in written form. In this way there will be a written record for future reference in case of any questions about the agreement. Edwards has documented this contextual information in a memo to Bruno Freitag. A copy of this memo appears on the following page.

MEMO

October xx, 19xx

TO: Bruno Freitag

FROM: Lionel Edwards

RE: Program Planning

As we discussed in our meeting on Wednesday of last week, Hamilton Unified School District's evaluation project is ready to enter the program planning phase. As I understand it, I will be helping the planning committee select a biology program to meet major objectives identified during the needs assessment.

The agreed upon budget for program planning activities is $xxxx. As we discussed, the following people will be on the planning committee:

Evert Walker, Principal, Hamilton High School
John Lanich, Biology Teacher
Linda Schroeder, Biology Teacher
Eric Steffansen, Biology Teacher
Maria Roberts, Biology Teacher

The projected date for the completion of program planning activities is January 1, 19XX, at which time the planning staff will have selected a biology program.

I will keep you informed about the committee's activities through short monthly reports.

Lionel Edwards

Program Designation

In this phase of the evaluation, Edward's task is to help program planners select a biology program to meet the high priority objectives identified during needs assessment. Edwards will suggest ways to fully utilize the people involved in the program planning and the most appropriate techniques for these people to employ in program selection.

Program Planners

.Consultants

The primary advantage of hiring a consultant is acquiring expertise which the planning committee lacks in a specific subject area. Employing consultants may also prove to be a financial asset since it is generally more economical to hire high level expertise for a short time than to employ someone full time who lacks necessary qualifications. Even though the consultant may merely substantiate the decisions of the planners, a second opinion from an outside source is valuable: the confidence gained from such support is a distinct asset. Furthermore, if there is not unanimity of opinion among the planners, the consultant may be useful in mediating the differences between committee factions.

Too often, however, the services provided by consultants are not used effectively, resulting in dissatisfaction with the consultant's performance. Employers can usually achieve a mutually satisfactory exchange of information if they follow the six guidelines listed below.

.*Specify in advance and in writing just what the consultant should do.* Write the job description, including any restrictions, such as "it must be done by January 1" or "it must involve all 10 principals in the district."

- *Send descriptive information about the school and the programs being considered to the consultant before work begins.* On large projects, it is often better to pay the consultant for reading about the project before arriving on site so that the consultant's visiting time can be used profitably.
- *Require and pay for a written report from the consultant describing the study, its rationale, and any recommendations.* These written reports will scare off many charlatans and provide a good record of what was said and why. Indicate the desired length and content of the report during negotiations with the consultant.
- *Schedule a meeting with the consultant both before and after the day of consultation to discuss the project.*
- *Provide the consultant with a list of names and job titles of the persons with whom the consultant will be working.* This procedure will produce an informality that often leads to a more productive atmosphere.
- *Find out if the consultant plans to use supporting help and who this will be.* It may be that the consultant does not really do the job alone, especially on jobs that take more than a day or two.

Rating Committees

When correctly used, the rating committee can be an effective communication method for program selection with advantages not found in other approaches.

Since it can include many people, a rating committee ensures that a number of viewpoints will be considered when selecting a program. In addition, people included on the rating committee usually develop a feeling of direct involvement in the program planning. Committee members should include staff directly involved in the program (e.g., teachers, administrators, aides); persons with curriculum expertise in the area; persons with evaluation expertise; representatives from the community, especially groups with a particular interest in the program (e.g., parents with children in the program, businessmen who hire graduates from the program, people in civic groups, students, etc.)

Disadvantage of the rating committee is that it often becomes an unwieldy tool through lack of a systematic approach to the committee's assignment: (1) the group's discussions get away from the subject at hand, (2) individual members of the committee monopolize the discussion, and (3) members do not stay within their areas of expertise.

Some suggestions for improving the effectiveness of planning committees include:

- *Specify the decision-making role of the committee at the beginning (e.g., advisory capacity to Director of Curriculum).*
- *Encourage the committee to specify their objectives and procedures so that it can approach its assignment in a systematic fashion*
- *Encourage the committee and the decision maker to agree on a timetable for program selection.*

.Techniques

On the next few pages some of the basic program selection techniques Edwards might recommend and the major advantages, uses, and limitations of these techniques are described.

The strategies presented do not represent all those available, nor do the descriptions of them pretend to be a "how-to-do-it-yourself" guide. This section will give you only an idea of the range of tools and techniques that an evaluator might recommend or develop. The techniques that are discussed, however, are ones that evaluators have found feasible and sensible for many kinds of planning activities.

The following approaches are discussed:

- Delphi Technique
- Planning Checklists
- Systems Approaches
  (with special attention paid to Decision Matrices)

Delphi Technique

The Delphi technique is a procedure for obtaining and organizing the views of experts. It is similar to the rating committee in that several people pool their knowledge and expertise to predict which will be the best program. The term derives from the Oracle of Delphi in ancient Greece who also made predictions. Unlike the rating committee, however, the Delphi approach is a structured dialogue between experts who may never meet each other face to face. The steps used to conduct a Delphi study are similar to the postal polling technique discussed in the Needs Assessment section of the workshop.

- *A questionnaire about various program alternatives (and the major aspects of them) that might fulfill important educational needs is developed.*

  This questionnaire might request suggestions for alternative ways of meeting these needs and/or ask for judgments about the efficacy of certain possible programs in terms of factors such as cost, feasibility, and likelihood of success. This questionnaire should also include space for the experts to specify the reason(s) for their judgments as this information may be needed later.

- *The questionnaires are then distributed to a panel of experts.*

  Since it is not necessary and often not desirable to have these experts meet, the questionnaire can be mailed to the people who agreed (probably for a fee) to serve on the Delphi panel.

- *When the questionnaires are returned, the results are tallied to determine the areas of agreement and disagreement.*

  For example, the experts might generally agree that a particular program is best, but might disagree as to the actual feasibility of implementing that program in the school.

- *When disagreement occurs, a second questionnaire containing the various reasons given by the experts for their initial judgments is distributed to the panel.*

  For example, the questionnaire should contain comments reflecting the divergent opinions on the feasibility of installing the program. These statements are purposely not identified by author. The experts examine these statements and re-evaluate the factors on which there was disagreement in the hopes of achieving consensus. They also include their reasons for any departure from their original evaluation.

- *This procedure can be repeated as often as necessary and feasible to achieve closer agreement, hopefully resulting in a better prediction.*

The Delphi technique is especially useful when the evaluation situation is so complex and time so limited that the only economical and feasible way of gathering the necessary information is to trust the opinions of experts. In addition, the Delphi approach eliminates the necessity and expense of bringing together experts from different parts of the country. Moreover, mailing out questionnaires to these paid consultants will increase the likelihood that all the desired views are expressed: it is more convenient for the experts to respond by mail than to meet at a prearranged time and place. Finally, since the experts interact solely on a written basis, the more wellknown or more persuasive panel members cannot cause a "band wagon" effect by stifling opinions or swaying views of others without evidence.

The disadvantages of this approach to selecting a program lie mostly in the cost of enlisting the experts' help, the difficulties of interacting by mail, and the possibility of making errors in judgment in selecting the participants. Further, there is no guarantee that the procedure will result in consensus, but it does provide a good climate in which people can "discuss" their views and change them if they think necessary without the ego problems associated with winning or losing an argument.

Program Planning Checklists

A planning checklist, another program selection tool, consists of a set of considerations for evaluating the merits of alternative programs or program components. The major purpose of a checklist is to ensure that all key factors are examined for each alternative. It is useful in identifying important facets of the program, such as equipment availability, which may otherwise be overlooked in considering the program's value. The chief advantage of the checklist, therefore, is in focusing the planners' attention on the critical considerations in evaluating any alternative.

One major limitation is the possibility of omitting important considerations from the checklists either from oversight or lack of agreement among program planners. Moreover, a checklist does not provide a systematic way for distinguishing program alternatives because it fails to assign any value or importance to the factors being considered. It only provides information concerning the presence or absence of key factors. In spite of these limitations, planners have found checklists quite useful in evaluating the merits of alternative programs, especially in isolating the absence of important factors.

The figure on the next page shows a planning checklist. It was used to examine the major program characteristics of three alternative courses. Even though Program C received more checks (indicating it contained more of the features the planning staff sought) this program was deficient in the area of instruction. Program B, with a very similar check total, lacked some of the characteristics in management, but was much stronger in instruction. The program planners, feeling that strength in instruction was more important than additional factors in management, selected Probram B for implementation.

MODEL PROGRAM PLANNING CHECKLIST

Program	A	B	C
A. General Characteristics			
.Do the program goals coincide with desired district goals?		✓	✓
.Is the program financially practical?	✓	✓	✓
.Does evidence demonstrate that the program will increase student achievement?		✓	✓
.Does the program use the desired instructional approach?		✓	✓
B. Internal Quality			
Instruction			
.Are the objectives stated measurably?	✓		
.Do the materials present instruction on the objectives?	✓	✓	
.Do the materials and procedures provide sufficient practice and feedback for each objective?		✓	
.Do the materials present accurate, quality information on the subject area?			
.Are the materials interesting?	✓	✓	✓
.Are the materials easy to use?	✓	✓	✓
Assessment			
.Are frequent assessment measures provided?	✓		
.Are end-of-program measures provided?	✓	✓	✓
.Are record keeping devices easy to use?	✓		✓
C. Management			
.Are training and orientation procedures provided?			✓
.Are administrator, teacher and student procedures clearly defined?			✓
.Will the program not interfere with other programs?			
.Will the success of the program positively affect student performance on other programs?	✓		
.Are the necessary personnel available to plan and run the program?	✓	✓	✓
.Is sufficient space available?	NO INFORMATION GIVEN		
.Are the required facilities, equipment and materials feasible?	✓	✓	✓
.Is the program feasible in terms of logistics, communication, budgeting and time requirements?			✓

Systems Approaches

Several different kinds of procedures fall under the heading "systems approach." The basic purpose of all systems approaches is to develop a systematic way of considering complex problems so that more enlightened (and hopefully better) decision making can take place. The five basic elements in a systems approach are as follows:

1. Objectives which are clearly defined
2. Alternatives for achieving each objective
3. Costs related to achieving the objectives
4. A model to display the relationship between objectives, alternatives and cost
5. Decision Rules to choose the best alternative

Some common systems approaches are Linear Programming Simulation, PPBS (Program Planning and Budgeting Systems), and Systems Analysis. Since each of these approaches has individual advantages and disadvantages, it would not be feasible to discuss them all in this workshop.

To clarify the use of systems approaches in program planning, let us examine one which provides valuable data to the decision maker and is easy to use. This approach involves the construction of a decision matrix.

Decision Matrix

Although there are several varieties of such matrices, they all have a common series of procedures for determining the relative merit of each program alternative on some numerical scale, such as 1 to 10. The obvious advantage of decision matrices is that they allow the planners to readily identify the relative overall value of each alternative.

Their major disadvantage is the number of necessary (and often arbitrary) decision rules required to construct them. This makes a decision matrix somewhat more difficult to construct than a program checklist but it provides much more information about alternatives. Since a decision matrix is much easier to use than some of the more highly sophisticated systems approaches that might be employed, planners often select such matrices over other systems approaches.

An example of a decision matrix used to compare three alternative programs appears on the following page. Note that like the program planning checklist, various program factors to consider in determining the relative merit of alternative programs (general characteristics, internal quality, management characteristics) are listed. To complete this decision matrix these steps were followed:

1. The planning staff assigned weights to each factor. A total point weighting system of 100 was used in the matrix. This means that the 100 points were divided among the factors in terms of importance. The factors they considered to be most important in the program were weighted most heavily. Major considerations such as goal agreement and instruction or objectives were assigned weights of 10.

Final Ratings of Alternative Courses

Program Consideration	Weight	Alternative Programs					
		Program A Rating*	Score	Program B Rating*	Score	Program C Rating*	Score
A. General Characteristics							
1. Goal Agreement	10	2	20	5	50	5	50
2. Cost of program material	7	5	35	4	28	1	7
3. Effectiveness	10	2	20	4	40	1	10
4. Instructional Approach	4	0	-	4	16	5	20
B. Internal Quality							
1. Instruction							
a. Measurable objectives	4	5	20	0	-	0	-
b. Instruction on objectives	10	3	30	5	50	0	-
c. Practice and feedback	6	2	12	5	30	3	18
d. Accuracy	4	4	16	5	20	4	16
e. Interesting	4	3	12	4	16	7	28
f. Ease of use	3	5	15	2	6	3	9
2. Assessment							
a. Frequent checks	6	5	30	0	-	0	-
b. End of program	2	5	10	3	6	5	10
3. Record Keeping	1	5	5	0	-	3	3
C. Management							
1. Training and orientation	1	0	-	0	-	4	4
2. Clear procedures	3	0	-	3	9	2	6
3. Other program interference	2	0	-	0	-	0	-
4. Positive student performance on other programs	2	2	4	0	-	0	-
5. Personnel availability	3	5	15	4	12	1	3
6. Space	6	0	-	0	-	0	-
7. Facilities, equipment and materials	6	4	24	3	18	5	30
8. Logistics, communication, and time	6	0	-	0	-	4	24
TOTAL	100		268		301		238
Final Average Rating			2.68		3.01		2.38

*Ratings: 5=Very Desirable; 4=Desirable; 3=Average; 2=Undesirable;
1=Very undesirable; 0=Not present

2. After weights were assigned, the planning staff used a 5 point scale (1=very undesirable, 5=very desirable) to rate each program on each factor. For example, program B included clearly stated measurable objectives, instruction that was relevant to all objectives, and a great deal of practice and feedback. Therefore the program was rated 5 for all these considerations.

3. Finally, the planners multiplied the rating given every characteristic of each program by its assigned weight to yield a score for each characteristic. For example, the score for goal agreement for Program C is 5x10=50 since 10 is the weighted importance at goal agreement and 5 (very desirable) is its rating. When all scores for each program were summed, a final average rating for each program was computed by dividing these totals by the total number of weights assigned (i.e., 100). As a result of this process the program staff selected program B.

A major advantage of the decision matrix is that it organizes information about alternatives in a way that facilitates choosing among them. When the number of alternatives and their complexity increases, it becomes increasingly difficult to remember the specific details of every alternative, and a systematic technique such as the decision matrix is useful.

One problem with the decision matrix is that the bases for determining weights and ratings often are not as solid as they should be. They tend to be based on subjective criteria and the judgment and bias of members of the planning staff. Also, as in the case of the planning checklists, neglecting to include important considerations in the construction of decision matrices will seriously impair the validity of the results. The use of these matrices may require the development of additional techniques for handling special problems, such as exceptionally high or low ratings on some considerations. It should be remembered that, despite these difficulties, the information obtained from decision matrices is generally better than that gathered by simply asking the opinions of experts or using a planning checklist. Decision matrices are more valuable because the alternatives are fully evaluated and ordered in terms of their utility. When needed, more sophisticated matrices may be constructed, such as those that consider the probable accuracy of the information on which the ratings are based and the confidence of the rater in making his judgments.

EXERCISE 3

Decision Matrix: Weighting Program Considerations

In this exercise you will assume the role of one of the members on the planning committee to select a biology program for Hamilton High. You have been involved in the construction of a decision matrix for selecting a biology program. You and your fellow committee members have already agreed on the various characteristics that should be considered. The next step will be to assign weights to these factors.

Assume the role of a biology teacher, evaluator, or curriculum coordinator. Work individually. Using a weighting system of 100 points, assign a weight of 10 or less to each characteristics according to how important you feel that factor should be when selecting a biology program for Hamilton High School. Enter the assigned weights in the column of your answer sheet labelled "initial weights."

Next compare the weights you assigned with those assigned by your team members. Discuss the assigned weights and compromise to arrive at a final rating system for your team. Enter these weights in the column labelled final "weights."

Program Considerations	Initial Weight	Final Weight
A. General Characteristics 1. Agreement of program goals with district goals		
2. Cost		
3. Proven effectiveness of program in other settings		
4. Program's uses instructional approach desired by district		
B. Instruction 1. Program has measurable objectives		
2. Program teaches to objectives		
3. Program provides practice and feedback		
4. Material is accurate, up to date		
5. Program is easy to use		
C. Assessment 1. Program includes frequent checks on student progress		
2. Program includes end of program test		
D. Management 1. Program requires little special technical training to implement		
2. Program provides clear procedures for teachers to follow in implementation		
3. Program has little potential for interfering with student achievement in other classes		
4. Availability of personnel to implement program		
5. Space requirements		

Biology Program Selected for Hamilton High

Once Edwards and the program planning committee agreed on the final weightings in the decision matrix, they were able to compare the programs being considered. The program they selected is described below in Edward's report.

Living Systems: An Introductory Biology Program

I. Overview

As a result of a needs assessment conducted in Hamilton Unified School District, the district determined that the biology curriculum was in great need of improvement. The present program fails to develop adequate knowledge of the subject and has adverse effects on student attitudes.

The major goals of the biology program, Living Systems, roughly parallel those of the district: (1) improve student attitude toward biology; (2) improve students' understanding of fundamental biological processes; (3) improve students' understanding and appreciation of their local environment.

II. Program Description

A. Instructional Characteristics

1. Program Goals

   .To improve student attitude toward and interest in biology

   .To improve students' understanding of the relationship between organisms and their environment

   .To improve students' understanding of the chemical bases of life

- .To improve students' understanding of reproduction, heredity, and genetics

- .To improve students' understanding of major classification of organisms

- .To improve students' understanding of the relation between the structure and function of organs

- .To improve students' understanding of man's responsibility contribution to the environment.

2. Program Objectives

- .Students will display a positive interest in and attitude toward biology by any of the following behaviors:

  (1) requests to use biology materials

  (2) voluntarily reads books on biology

  (3) does biology activities and experiment at home

  (4) volunteers to bring in materials

  (5) volunteers to do research

- .Students will demonstrate their understanding of the relationship between organisms and their environment by explaining various interactions between organisms in a given environment (e.g., food chain, population dynamics and changes, food production.)

- .When presented with a chemical process necessary for life students will explain and/or demonstrate this process (e.g., enzyme active, ATP cycle, fermentation, respiration, photosynthesis)

- .Students will correctly describe specific processes involved in reproduction and tell how they contribute to the continuity of life (e.g., mitosis, meiosis, DNA replication; Mendelian heredity; natural selection; asexual reproduction)

- .Presented with a drawing of an organism representing a certain phylum, students will identify the phylum to which it belongs and list the characteristics of the organism that are typical of that phylum.

- .Presented with an organ or organ system students will describe the structure of the organ or organ system and relate it to the functions it performs (e.g., blood vessels, muscles, fruits and seeds, small intestines)
- .When presented with sources of pollution in the local environment, students will list strategies which could result in decreased pollution.

B. Administrative Characteristics

1. Pupils: The program is designed for 10th grade high school students. Approximately 120 students will participate in the program in September in Hamilton High School. Four classes will be offered, with approximately 30 students in each class.

2. Resources: The program includes:

.Teachers Manual

.Laboratory Manuals

.Supplementary Materials:

(i) Practice Kit

.contains individual cards. Each card gives more practice on all concepts in Living Systems.

(ii) Instruction Kit

.contains individual cards on each concept taught in Living Systems. Each card includes instruction and practice simplified from that found in the textbook.

.Testing Materials:

-ditto masters for chapter pre and post quizzes

-unit tests

-midterm and final exams

3. Special equipment: microscopes, dissecting kits, specimens

4. Schedule:

   .Biology classes will meet every day for approximately 50 minutes per day.

   .3 hour laboratory session is scheduled weekly.

   .Two classes will go on monthly field trips while 2 will go on field trips twice a month.

   .Half the students requiring remediation will use the Instruction Kit, half will use the Practice Kit.

SECTION D OUTLINE

PROGRAM PLANNING TOOLS AND PROCEDURES

I. Selecting Program Planners

  A. Hire an outside consultant

    1. Advantages

      a. Provides high level expertise in a specific subject area

      b. Provides opinion from an outside source

      c. May fulfill need for mediator

    2. Disadvantage: services often not used effectively

    3. Guidelines for effective use

      a. Write job description

      b. Provide pertinent information to the consultant in advance

      c. Require written report

      d. Organize meetings before and after the consultation

  B. Organize a rating committee

    1. Advantages

      a. Facilitates acceptance of the program because members (including persons who will participate in the program) feel directly involved

      b. Allows for consideration of various viewpoints

    2. Disadvantage: committee can be unwieldy

    3. Guidelines to ensure effectiveness of committee activity

      a. Specify decision-making role

      b. State objectives

      c. Define time lines

      d. Appoint chairman with skills in group leadership

II. Techniques for Planning Programs

A. Conduct Delphi study

1. Description: consensus among panel of experts who interact by mail through an intermediary

2. Procedure

    a. Develop questionnaire about program alternatives

    b. Distribute questionnaire by mail to panel of experts

    c. Analyze results to determine areas of agreement and disagreement

    d. Resolve differences through follow-up questionnaires until reasonable consensus is reached

3. Advantages

    a. Provides expert opinion when data are lacking

    b. Provides feasible way to gather information in a short time, without the logistics problems inherent in organizing a meeting

    c. Provides non-threatening climate for each person to express his views as fully as he desires

4. Disadvantages

    a. Cost

    b. Difficulties in mail interaction

    c. Difficulty in selecting appropriate experts

    d. Possibility that a consensus will never be reached

B. Construct planning checklists

1. Description: list of all considerations to be evaluated in each alternative plan

2. Advantage: states all key factors and focuses attention on them

3. Disadvantages

    a. Items can easily be omitted

    b. No system for organizing the information collected

C. Use systems approaches

1. Description: a system to study program considerations so that the information generated can be used to select programs

2. Procedure

   a. State objectives

   b. State alternative methods for achieving objectives

   c. Detail costs

   d. Construct model that relates objectives, alternatives, and costs

   e. Develop a decision rule for program selection

3. Example: decision matrix

   a. Procedure

      (1) Specify alternatives

      (2) List and define all program considerations relevant to program selection

      (3) Assign weights to program considerations

      (4) Rate each consideration within each program alternative

      (5) Multiply weight and rating for score

      (6) Add all scores for each alternative; compare totals for final decision making

   b. Advantage: systematic way to identify relative overall value

   c. Disadvantages

      (1) Assignment of weights and ratings, which is the basis of the whole system, is often arbitrary

      (2) Failure to include important considerations will seriously impair validity of results

      (3) Special techniques may be required to handle very high or very low ratings

SECTION E

EVALUATION PLANNING

PLANNING THE EVALUATION
E. Evaluation Planning
  Purpose of the Evaluation:
    Formative and Summative
  Context Considerations
  Evaluation Questions
F. Measurement Instruments
G. Evaluation Designs

Since Lionel Edwards has been working with the planning staff throughout program planning, he will now work with the staff to plan the evaluation of the program which is to be implemented in the fall. Such early evaluation planning is desirable but not always possible. The evaluator is frequently called in after the program has been implemented or, in some cases, even after it has been completed. Evaluation activities will include:

.identifying the purpose of the evaluation
.formulating evaluation questions
.specifying measurement instruments
.selecting the evaluation design
.planning the analysis activities

The first two areas are treated in this section of the workshop.

Purpose of the Evaluation: Formative and Summative

Clearly the intended function of the evaluation will control the structure of the evaluation plan. In evaluation literature, a distinction is made between two kinds of evaluation: formative and summative. Formative evaluation provides data that will allow decisions to modify a program; summative evaluation influences decisions to continue, terminate or expand a program. These differing purposes in turn imply different evaluation procedures. The type of evaluation affects a number of aspects of the evaluation. Table 3 summarizes these dimensions for formative and summative evaluation. As you can see, the focus of the evaluation will affect several dimensions including design, data collection, analysis, the evaluator's role and the reporting strategy.

TABLE 3

Characteristics of Formative and Summative Evaluation

Evaluation Dimension	Formative Evaluation	Summative Evaluation
1. Purpose	1. Program Improvement	1. Program expansion, continuation
2. Stage of program Development	2. During new program or program being used for the first time	2. Established program used in a district for a period of time
3. Program Characteristics	3. Early draft; changeable during progress of program	3. No changes; fixed
4. Focus	4. Usually one program	4. May compare two or more programs
5. Data strategy	5. Intensive, specific (measure enroute skills, interim components, terminal skills); clinical	5. Less intensive; more general (measure overall program, and terminal skills); objective
6. Instruments	6. Many, very detailed. Intensive process analysis	6. Few; on major program objectives, less process analyses
7. Design	7. To assess effectiveness within the program: Case Study; Pretest-Post test;	7. To assess effectiveness within and between programs: all designs
8. Data collection	8. Frequent (at beginning during and after program)	8. Less frequently (usually at beginning and always at end of program)
9. Reporting	9. Immediately to program planners and staff; detailed; formal and informal modes	9. At end of evaluation. Given to administrators, funding agency
10. Evaluator's role	10. Nurturing; may be internally involved with program	10. Objective; 3rd party; may need to be external to the program

FORMATIVE EVALUATION

Purpose

Since the purpose of formative evaluation is to improve a program, the evaluation is usually conducted for a program just developed or newly implemented in a school or district. The effectiveness of program activities and materials has either not been validated at all, or have not been verified as appropriate for the context and students of a particular district. The program, therefore, can be changed, and improved or adapted to the district's situation as the program is being used. Generally the evaluation will focus on the refinement of this one program rather than attempting to compare it to other programs. Fairness and logic imply that a new program should be allowed a "debugging" period before judgements about its merit are made. Methodologically the assumptions of experimental design and analysis make it difficult or impossible to compare a continually changing program with an established one.

General Strategy

The general strategy of formative evaluation is to collect intensive specific information about the feasibility and effectiveness of various program processes. A diagnostic, clinical approach may be adopted in order to distinguish effective procedures and materials from those which require modification, improvement or elimination. Measurement instruments generally include items on all program objectives, entry, enroute and terminal. A variety of process measures can identify and cross validate process effectiveness.

Evaluation Design

The evaluation design usually involves a case study approach to monitor program implementation and progress. It may also include experimental variations within the program to contrast the effectiveness of alternative activities or materials. To permit continuous program diagnosis and modification, data collection is frequent, occurring before, during (preferably after each program unit) and at the end of the program.

Reporting Schedule

Ongoing program improvement also requires frequently reporting data to program staff. These feedback trouble shooting reports can be informal in contrast to the formal written reports which may be required by a funding agency.

Evaluator's Role

To accomplish a trouble shooting mission, the evaluator must be perceived in a nurturing supportive role. The evaluator may be a member of the program staff.

SUMMATIVE EVALUATION

Purpose

Given that the purpose of summative evaluation is to support decisions to continue, terminate, or expand a program, the evaluation should be conducted for an established program whose activities and materials are fixed. The summative evaluation may assess the overall effectiveness of just the targeted program, or better yet. may compare that program with other, competing programs.

General Strategy

Since the interest is in overall effectiveness, the data strategy is less intensive and specific than in formative evaluation. Interest is in documenting major program activities and in measuring performance on major program objectives. Data are usually collected less frequently, only sometimes on unit tests as in formative, but definitely on program pre and post tests. Wherever possible, the program's effectiveness is measured by comparing achievement of objectives by students in the new program with achievement of objectives by students in similar programs or with students not enrolled in any program teaching the target objectives.

Evaluation Design

The evaluation design may be any of the single program or comparative experimental designs.

Reporting Schedule

Reports tend to be primarily formal; submitted perhaps once or twice during the evaluation but usually at the end of a program. The main audiences for the reports are usually the funding agency and district level administrators, rather than the program staff.

EVALUATOR'S ROLE

The evaluator must often adopt an objective, third party role and may be required to be external to the program to avoid charges of bias.

In theory, formative and summative evaluation vary according to the several dimensions just described. In practice, however, evaluations may occur along a continuum ranging from formative to summative purposes and procedures. Clearly the most effective evaluation will be that which can provide the most useful, high quality data so that informed decisions can be made.

Therefore, before proceeding with the evaluation plan, the evaluator should find out if decision makers are interested in a formative or summative evaluation.

Frietag has called Edwards in to discuss the evaluation with him and with Sandi Silverthorn, Curriculum Specialist. Part of their conversation is transcribed below:

Conversation among Frietag (F), Edwards (E), & Sandi Silverthorn(s)

E. Sandi, Bruno, how are you?

F. Fine, Lionel, glad you could make it. I had a heck of a time getting through that construction area.

S. Yeah, me too. They're always tearing up those streets and obstructing traffic.

(Secretary comes in with coffee which he distributes to the people)

E. So, you mentioned over the telephone that the school board is happy with the program and we can start planning the evaluation for next year. Generally speaking, what type of information are you looking for?

F. How about finding out how the new program is running and how students do at the end of the year?

S. Oh yes, and how it compares to the traditional program.

E. Those are all important concerns which we'll want to answer. However, we probably don't want to do it all at once. First we should evaluate the program in progress and look at the data to see if program modifications are necessary. Then, when all is running smoothly we can evaluate the program again and make decisions about its effectiveness by comparing it to, say the traditional program. This procedure will enable informed decisions to be made about continuing, terminating or expanding the program.

F. So, they're actually 2 separate evaluations.

E. That's right. They should be. Since the program is a one semester course and we only have a year for this evaluation, I suggest that we do the formative evaluation the first semester and the summative the second.

F. Sure, that sounds fine. We have six biology classes so we'll have six opportunities to observe the effects of the new program -- Living Systems.

S. Well, I don't know. What if this program doesn't work out for some reason? Do we want all of the classes involved in it?

E. Good point. Probably the best thing to do would be to have four classes take the new program the first semester and two not. Again, the second semester, have two classes keep the traditional program and four adopt the new one which hopefully we will have improved after its first tryout. It really would be nice to have another improvement period for the second semester - but we only have the year to complete both evaluations. So in the summative evaluation the second semester we will compare the new program to the old one.

F. That sounds great.

(FADE OUT)

Context considerations

Edwards, Silverthorn, and Frietag have come to an agreement about the kind of evaluation to be planned for the biology program and the time frame in which to do it. They have agreed on a formative evaluation during the first semester, summative the second semester. The information Edwards must now gather includes:

- *What are the purposes of the evaluation?*
    - specific decisions to be made, questions to be answered?
- *What resources are available for the evaluation?*
    - time and money allotted for the evaluation
    - personnel who will involved in the evaluation
- *What kinds of evidence of program success is requested, possible, and technically appropriate?*
- *What types of audiences are there for information about the program and the evaluation?*
    - audiences for interim and final reports
    - reporting deadlines

In addition to questions about the evaluation, an evaluator who is called in to do an evaluation at this stage would need to gather background information concerning the school and community structure. Since Edwards has been involved with the Hamilton District since it's needs assessment, he knows a great deal about the program to be evaluated. But if he were to be called in now, he would have to become fully acquainted with the program's characteristics:

- .subject matter
- .program's origin
    - how it got started
    - prime movers
- .general goals
- .how they were established
    - (needs assessment, mandated, other)
- .specific objectives
- .how they were selected
    - (research, theory, rationale)
- .students served by program
    - (age, grade level, ability level)
- .program and instructional strategies

In this workshop's simulation, these program characteristics have been summarized in the program description on pp. D-20-D-23.

EVALUATION QUESTIONS

The next step in planning the evaluation is to specify the evaluation questions it will answer. Whether an evaluator conducts a formative or summative evaluation, it is advisable to organize the evaluation around a set of evaluation questions. These questions should address each of the concerns such as outcomes, costs and the like, that program planners have about a program.

Evaluation questions should be arrived at jointly by the evaluator and planning staff. They can be placed into the following broad categories which, although somewhat arbitrary, will help organize the questions that are formulated:

.*OUTCOME QUESTIONS -- Is the program meeting its objectives?*

Outcome questions focus on the attainment of psychomotor, cognitive, and affective en-route and end of program objectives

.*IMPLEMENTATION QUESTIONS -- Is the program being implemented as planned? What does the program look like in operation?*

Implementation questions focus on program processes, both instructional and administrative

.*COST QUESTIONS -- What is the cost of running the program including dollar expenditures, time, resources?*

Cost questions can focus on many different aspects of a program including: (1) dollars spent on teachers salaries, materials etc; (2) wear and tear on equipment; (3) allocation of personnel

.*RESEARCH QUESTIONS -- How do specific program variables relate to program outcomes?*

Research questions involve a planned variation in program processes to determine what impact, if any, these variations will have on program outcomes

.*UNANTICIPATED OUTCOME QUESTIONS -- What other effects, either positive or negative, can be attributed to the program?*

Unanticipated outcome questions are designed to make evaluators and planners aware of the importance of attending to unplanned program effects. It is very probable that the program will have effects that were not planned but which nevertheless have great impact. Of course, unanticipated outcome questions cannot be clearly formulated at the beginning of an evaluation. However the evaluator should be on the look out for unanticipated outcomes, and design or select instruments that can might be sensitive to them.

.FORMATIVE/SUMMATIVE CONSIDERATIONS FOR EVALUATION QUESTIONS

As mentioned at the beginning of this section, the intended purpose of the evaluation will control many components in the evaluation plan. One of these components is the evaluation questions which will be answered. While all five categories of evaluation questions are asked for both formative and summative evaluations, the focus of evaluation questions and the frequency with which questions in certain categories are asked varies according to the type of evaluation being conducted.

Outcome questions in formative evaluation usually include questions on the attainment of enroute program objectives by students in a newly implemented program. In summative evaluation the questions will frequently compare the performance of students in one program on major program and interim objectives with the performance of students in another program.

Implementation questions are emphasized during formative evaluation. Instructional sequencing and material usage are carefully monitored and several administrative and instructional variables are examined. In contrast, implementation questions in summative evaluation focus only on key process variables, both administrative and instructional. The primary reason for collecting implementation data in summative evaluations is to more completely characterize a program and explain program outcomes, anticipated and unanticipated.

Cost questions in formative evaluation often focus on the costs associated with the planning and start up phases of a project. The costs associated with installing the program are of prime interest. In summative evaluation cost questions focus on costs during the operational phase of a fully developed program: personnel, space, equipment, and supplies. These cost questions can also be comparative in nature, comparing the cost effectiveness of competing programs.

Research questions are most appropriate for formative evaluation since the answers to these questions will likely be incorporated in any modifications to be made in the program. However, research questions may also be asked in a summative evaluation if a formative evaluation has not been conducted and the program staff is especially interested in investigating the effects of varying an instructional and/or administrative component of the program.

The planning committee and Edwards have drawn up a list of evaluation questions for the formative evaluation of the biology program. They are listed below by category:

Outcome

1. Do students show a positive attitude toward and interest in biology?

2. Are students able to explain various interactions between organisms and their environment?

3. Are students able to explain or demonstrate various chemical processes necessary for life?

4. Are students able to classify organisms according to phyla and list the characteristics of various phyla?

5. Are students able to describe the structure of an organism and state how its structure relates to its function?

6. Are students able to list strategies to minimize environmental pollution?

7. Are students able to correctly describe specific processes in reproduction and tell how they contribute to the continuity of life.

Implementation

1. Are the field trips being conducted as planned?
2. *How much* subject matter content is being covered by individual teachers?
3. Are teachers following the programs planned instructional sequence?

Cost

1. Is the performance of students using the Practice Kit or the Instruction Kit significantly enhanced relative to the cost of the Kit?
2. What is the cost-effectiveness of field trips?
3. What is the cost of wear and tear on laboratory equipment?

Research

1. Which remedial technique included in the Living Systems program, Instructional Kit or Practice Kit, is more effective in improving student performance?

2. What is the effect of the frequency of field trips on student attitude?

Unanticipated Outcome

.What other outcomes can be attributed to the biology program?

If planners want to pick up potential side effects, items designed to elicit them can be built into the measurement instruments. Unexpected "spin-offs" will hopefully be picked up by open-ended questionnaire or interview items, or in observations.

SECTION F

Section F

MEASUREMENT INSTRUMENTS

PLANNING THE EVALUATION

E. Evaluation Planning
F. Measurement Instruments
    Guidelines
    Paper and Pencil Test
    Reports
    Observations
    Records
    Formative/Summative Considerations
    Exercise
G. Evaluation Designs

Now that the planning committee, with the help of Edwards, has formulated a set of evaluation questions, Edwards and the program staff must begin to plan *what information* and *types of instruments* will be appropriate to answer these questions. The instruments may be selected, constructed, or adapted. Actual construction or adaptation of any instrument, however, will be influenced by the evaluation design Edwards selects. The design, which outlines what sample will be measured and when a measurement instrument will be administered, will affect the specifics of the measure (e.g. length of test, number of items per objective). These design concerns are discussed in a later section of the workshop. The initial task, the selection of the devices, is treated in this section.

Guidelines for Selection or Construction of Measures

The technical quality of the measurement instruments an evaluator uses underly the credibility of the evaluation results. The quality and utility of the data an evaluation yields are largely dependent on the measures used. There are several general guidelines the evaluator should follow when choosing or developing measures.

.Measures must meet the information needs of decision makers. Measures must be directly linked to the evaluation questions. The evaluator should critically examine all measures, especially published instruments, to be certain that there is a match between the measure and the decision makers' information needs.

.Measures must be appropriate for the project being evaluated. The measure must be at a suitable level of difficulty and match the ability level of program participants and others taking the measures.

.Measurement instruments should be as valid and reliable as possible.

The *reliability* of a measure refers to the consistency with which it assesses whatever it is intended to measure. If a test of reading comprehension has demomstrated a high degree of reliability, the evaluator can be reasonably sure that it will measure reading comprehension consistently with different groups and at different times.

The *validity* of a measure indicates the degree to which it measures what it says it measures. If an evaluator finds a test which is supposed to measure reading comprehension, it is important that information concerning the test's validity be examined to see if the test really does measure reading comprehension and not, for example, vocabulary.

To properly assess the technical quality of a published measure, the evaluator must examine the information concerning the reliability and validity of the measures that is usually included in the test manual. Similarly, if tests are developed by local school personnel, their reliability and validity should be determined. A consultant or person in the district trained in measurement methodology should be contacted to help in this task. The concepts of reliability and validity can take on relatively complex technical and statistical meanings when applied to test construction. When in doubt, the evaluator should call on a measurement specialist for assistance.

.Measures should be sensitive to the positive and planned as well as the negative and unplanned outcomes of a program.

.Data collection should be as minimally disruptive as possible. All measures should be implemented in such a way that they do not intrude excessively in the program's operation.

.Multiple measurement strategies should be utilized. Since evaluations are concerned with measuring a full range of program characteristics including activities, materials, costs, attitudes, and achievement, a range of measurement instruments is needed.

This section highlights major kinds of measures and discusses their advantages, limitations and general considerations for their construction.

.Paper and Pencil Tests
    .Test Formats and Construction
    .Norm Referenced Tests
    .Criterion Referenced Tests

- .Reports
    - .Questionnaire/Survey
    - .Interview
    - .Ratings
- .Observations
- .Records

PAPER AND PENCIL TESTS

Paper and pencil tests are the most common type of measure used in educational evaluation. Among the various types of tests available to the evaluator are standardized or norm referenced tests and criterion referenced tests. Tests also come in various formats--involving selected response items, and constructed response items.

.Test Format

Selected response or closed items supply the student with a question or incomplete sentence "stem" plus a list of possible answers--some correct plus some distractors. True-false, matching and multiple choice are three types of selected response items. Scoring merely involves tallying responses.

Open ended or constructed response items require a student to recall or reconstruct an answer with a minimum of prompting. Such items are relatively easy to construct but are more difficult to score than closed response items. Among the several types of constructed response items are essays, short answer, and completion items. Another type of constructed response format involves work samples. The student produces a product--paints a picture, sews a shirt, completes a job application form, writes a business letter, etc.

Test Construction

All tests should be written from a construction plan or a set of test specifications. In general, a test specification is a clear definition of the behaviors to be measured and the content area in which they will be measured.

The precise mode of specification varies according to the specific type of test to be developed. In the discussion of NRTs and CRTs that follows, one mode of specification for each of these types of tests is briefly discussed. More specific information pertaining to test construction is contained in the references at the end of the notebook.

.Norm Referenced Tests

Norm referenced tests are tests in which an individual's performance is compared to that of a large group of people--a "norm" group. Therefore, a student's score is reported in terms of his or her relative standing in comparison to the norming sample. For example, a student's score may be reported as falling at the 85 percentile which means that he scored above 85% of the people in the norm group. Scores may also be expressed as grade equivalent scores or standard scores. For more explanation of these methods of reporting scores you may refer to any introductory text on educational and psychological testing.

Norm referenced tests not only provide a relative score for each student, they also describe a student's performance in a general skill area such as reading comprehension, or mathematical reasoning. If a school board wants to know how the reading comprehension of fourth graders in their district compares with the performance of 4th graders throughout the state, a norm referenced test would provide these data.

Since norm referenced tests have been prominent measurement devices for many years, there is a rich technology of norm referenced test construction. The tests are designed to discriminate between learners; the items on the test are specifically designed so that only about half of the students taking any item answer it correctly.

Before using a norm referenced test the evaluator should check to see if the group on which the test was normed is comparable to the group which is to be tested. If the norming sample is different, either a different test should be selected or certain adjustments should be made when interpreting test results.

.Criterion Referenced Tests

Criterion referenced tests are tests in which an individual's performance is referenced to clearly specified skills, knowledge, or abilities. When mastery of specific skills or knowledge is to be inferred from test performance it is most appropriate to use a criterion referenced test. The results of criterion referenced tests show an individual's level of achievement for each area assessed in the test. Scores are typically expressed as percentage of items correct.

Published criterion referenced tests are available or they may be developed locally. Criterion referenced tests should be constructed on the basis of a clear set of rules. These rules may be described as defining a domain (i.e., specified content and performance which the test measures). The rules for the domain specify the content on which the test items are drawn and describe the observable behavior the student will be expected to demonstrate.

REPORTS

Under the general heading of reports fall questionnaires, interviews, ratings, and observations. To ensure that these measures yield the maximum amount of information, they should be developed in a systematic fashion. An evaluator should follow a general development cycle like the following:

.Construction of Reports

.meet with a small group of respondents selected from the entire pool of respondents (e.g. teachers, students, administrators, parents in a particular school district) to determine the focus and content of the instrument:

1) objectives to be addressed
2) use of open-ended, closed-response mode or a combination

In a closed answer the respondent indicates an already provided response. Multiple choice or checklists are common closed response forms. The response may be written (e.g. interview), or exhibited (e.g. observation behaviors). Recording answers merely involves tallying response categories chosen. The closed response mode if often used because of this ease in scoring responses.

An open-ended response mode, however, allows for a wider range of responses. Questions calling for respondents opinions and comments are often open ended. Information can be picked up that might not have been obtained if the respondent could only answer in a predetermined way. A disadvantage of this technique is the larger amount of time necessary to categorize and score the responses.

.develop a rough draft of the measurement device.

.pilot test the measure with a small group chosen from the respondent pool.

.pilot test scoring procedures, particularly of any open ended questions.

.revise the measure.

The last two steps may be repeated one or more times until the evaluator feels the measure will yield the desired information.

Questionnaires/Surveys

Advantages. Questionnaires and surveys are excellent measures to use to collect information from a wide audience. A major advantage of questionnaires is that they are an economical way to collect information from many people since they can be mass produced and mailed to various groups -- parents, teachers, students, etc.

Disadvantages. One of the drawbacks of questionnaires is that it is fairly difficult to construct a good questionnaire--one that will yield needed information and one that people will be willing to take the time to complete. Whenever it is possible, the evaluator should try to make the questionnaire as short as possible because people are more inclined to complete a short 1-2 page instrument than a long one.

Construction/Format. Questionnaire development should follow the steps outlined on the previous page. It is especially important to pilot test a questionnaire to ensure that it is unambiguous and asks the "right questions" to gather information decision makers need.

Interviews

Advantages. Interviews, an alternative to questionnaires, are especially useful with small groups of respondents. One advantage of interviews is that they enable the evaluator to collect more complete information than can be obtained using questionnaires. The interviewer is able to probe responses and ask follow-up questions about specific areas.

Disadvantages. A disadvantage of interviews is that they are extremely time consuming to conduct. In larger scale evaluations the evaluator will also have to invest additional time to train interviewers.

Construction/Format. Interviews can be structured or unstructured. Structured interviews utilize a type of closed response mode in that specific questions are arranged in a specific order. Neither the questions nor their order are varied from interview to interview. An unstructured interview, in contrast, consists of a few general questions with predetermined probes. The questions in an unstructured interview do not have to be asked in any set order.

It takes considerable time and effort to develop interview forms. Their development should proceed systematically as described on page F-7.

OBSERVATIONS

Construction/Format. Perhaps the best way for an evaluator to collect program implementation information and get in touch with a program is to conduct observations. Such observations can be informal and open, or formal and closed. Observations involve a casual look at a program by one or more observers followed by a report. They are quick and easy to conduct and are especially useful im small scale evaluations. Formal structured or closed observations follow a more strictly controlled methodology, specifying times of observation and methods for recording information. Observers enter the classroom at a preselected "typical" time, stay for a defined length of time, and are usually guided in their observations by an observation form. Formal observations are useful in larger scale evaluations. Generally, the information they yield, especially if observers are trained, is more reliable than that gathered via informal observation.

Advantages. A major advantage of observations is that actual instructional events can be recorded. Figure 4 below contains a portion of an observation checklist that was used to assess student attention span for an educational program.

Figure 4

OBSERVATION CHECKLIST
SHOWING INCIDENTS OF NON-ATTENTION

CLASS ACTIVITIES	Staring Out Window	Talking to One Another	Playing with Object	Staring Around Room	Dozing	
Announcements		III			I	
Discussion	I					
Lecture		I	I			
Individual Study		~~IIII~~				
Study Groups				II		

The observations were tallied, then the tally marks were counted and analyzed to determine the parts of the program to which students did or did not pay attention. Such observation checklists usually describe the behaviors to be observed in such terms as number of responses per minute, (e.g. the number of responses made by individuals versus the entire class as a group.) This kind of information is useful in deciding which instructional sequences and procedures are working well and which parts need improvement.

Disadvantages. A major limitation of observations is that their obtrusiveness may interfere with the working of the program, making teachers anxious and affecting the routine operation of the activities which are being observed.

There is some interest in the use of unobtrusive observation methods to avoid this problem. For example, the evaluator might examine samples of behavior and pieces of conversation in such varied settings as assemblies, faculty and student luncheons, playgrounds, faculty meetings, PTA meetings, and recesses to get a sense of teachers', students' and parents' reaction to a program.

Like interviewers, observers must be trained. Also, observations are time consuming since several observations may be necessary to provide reliable information.

RECORDS

There are many different records and reports which are available to the evaluator. They are one of the least expensive ways of obtaining information about a program although they should be used in conjunction with other measures to substantiate the information.

Many records that may be useful to the evaluator are kept as a matter of school routine: attendance records, tardiness records, report cards, folders, health records, etc.

Other records might be purposely kept for evaluation purposes: records of number and type of assignments computed, or teachers' logs of program objectives attained. In addition, the evaluator may wish to examine teacher's lesson plans.

Formative/Summative Considerations for Measurement Devices

As discussed earlier in the section on formative and summative evaluations, the focus of an evaluation has great impact on the evaluation plan including measurement selection and construction.

Since formative evaluation frequently documents program processes and outcomes, measurement devices are rather specific, focusing on progress in attaining all important instructional objectives (enroute, terminal), use of instructional materials, sequence and pacing of activities, etc. The emphasis of all instruments is diagnostic and, to collect information needed for program modification, a greater number of measures are administered more frequently than during summative evaluation. Appropriate measurement devices and strategies include unit tests, detailed interview schedules, frequent classroom observations.

In contrast, the measures used in summative evaluation are broader in focus. Observations are carried out less frequently (e.g. once or twice a semester). Tests may be administered at the beginning and end of the program and at one or two times during the year. Questionnaires may be extensively used while interviewing would likely be restricted to a small group of program participants.

EXERCISE 4

Measurement Devices

In this exercise you will find a letter Lionel Edwards has written to Susan Lightner. The letter lists some of the evaluation questions (one from each category). The column showing which measures will be used to answer the evaluation questions has not been filled in, but the audience for each measure has been indicated.

Working as a team, decide what measure or measures would be most appropriate for Edwards to administer to each audience or source. Each team member should assume one of the following roles:

- superintendent
- evaluator
- teacher
- curriculum specialist.

EXERCISE 4

Measurement Devices

Date

Dear Susan:

I am sending you a list of the evaluation questions and the measures I will be using to answer these questions. Also included is a rational for my decision. What do you think? Any suggestions or comments would be appreciated.

Evaluation Question	Audience	Measure
1. **Outcome**: Do students show a positive attitude toward and interest in biology as evidenced by: (1) requests to use biology materials (2) voluntarily reads books on biology (3) does biology activities and experiments at home (4) volunteers to bring in materials (5) volunteers to do research	Teachers Students	
2. **Implementation**: How much subject matter content is being covered by individual teachers?	Teachers	
3. **Cost**: What is the cost of wear and tear on laboratory equipment?	District purchasing department Lab Teachers	
4. **Research**: Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit, is more effective in improving student performance?	Students	
5. **Unanticipated Outcome**: Have any logistical problems been caused by the field trips?	District accounting office Teachers	

OUTLINE

MEASUREMENT DEVICES

I. Paper and Pencil Tests

  A. Norm-referenced tests: Examinee's performance on test items is compared with performance of other examinees

  B. Criterion-referenced tests: performance on specific items is compared against a stated performance standard

  C. Advantages

    (1) Superior quality, quantity, and scope of data

    (2) Ease of data collection and processing

    (3) Low cost for data of high quality

    (4) Reusability of instruments

II. Reports

  A. Questionnaires

    .Economical way to collect information from many people

  B. Interviews

    .useful with small groups of respondents

    .allows probing of responses

III. Observations

  .Checklist

    (1) Behaviors to be observed are listed

    (2) Occurrences of behaviors are tallied

    (3) Success of program components relating to listed behavior is evaluated

IV. Records

  .show attendance, student achievement, etc.

  .major advantage: inexpensive data source.

SECTION G

Section G

EVALUATION DESIGNS

PLANNING THE EVALUATION

- E. Evaluation Planning
- F. Measurement Instruments
- G. Evaluation Designs
  - Who
    - Sampling
  - When
    - Measures
    - Comparison groups
    - Randomization
    - Pretest
  - Formative Considerations
  - Summative Considerations
- H. Data Collection
- I. Analysis
- J. Reporting

At this point in the evaluation, Edwards has helped the planning staff formulate evaluation questions and then choose measures to obtain information about these questions. Another element that must now be included in the evaluation plan is the design. A design is a data collection plan which includes the who and when of the measurement devices.

The "who" are those persons completing a measure such as a test or questionnaire, or acting as the subjects of a measure, such as an interview or observation. "When" is the time plan for administration of the measure. This section of the workshop will discuss the who and when of the measure, design terminology, some commonly used evaluation designs, and particular design considerations for formative and summative evaluations.

"WHO"

When the evaluator and decision makers decide what measure will be used to gather information on an evaluation question, a respondent group will probably have been mentioned at the same time. For example, in the present evaluation, one evaluation question is: "Do students show a positive attitude toward and interest in biology?". Not only was an interview discussed as one of the appropriate measures, but an interview with teachers and students was stipulated at the same time. However, which specific teachers and students would be interviewed was not necessarily decided in that meeting. Now that Edwards is working out the design, it is time to specify who in the respondent group will be measured (i.e. which teachers and students).

SAMPLING. Usually it is not necessary or feasible to measure everyone in the population of the respondent group. Often there is not enough time, money, or personnel to cover such a large task. Therefore, some method of sampling is often employed. Sampling is a process used to select which members of the population will be given the measure. In sampling procedures, these persons are chosen in a systematic way so the sample is representative of the population from which it was drawn. This allows the findings from the sample to be generalized to the whole population. There are several commonly used sampling procedures:

'Random Sample. This technique was mentioned earlier in the Needs Assessment section. In this procedure each person in the population has an equal chance of being included in the sample. There are several methods to accomplish this equality. For example, each person's name could be written on a piece of paper and, as in a lottery, all the pieces thoroughly mixed and the appropriate number of names drawn. Since tearing up hundreds of little pieces of paper can be a prohibitive and messy task, however, tables of random numbers (found with directions in most statistical books) are often used.

•Stratified Sampling. If it appears that certain population factors (e.g. aptitude, sex, age) might have an effect on the measure, stratified sampling allows some control over this variability. In this technique, the population is divided into catagories or strata according to whatever factors one wishes to control. People are then randomly selected from each category. The number chosen from a category is proportional to the number found in the total population. This increases the likelihood that the sample will be representative of the population. For example, an evaluator wished to assess student attitude on a new program but thought that achievement level in the program might be a factor that would affect attitude. Therefore, she decided to give the questionnaire to a stratified sample that controlled for achievement. She divided the students into High, Medium, Low achievement groups based on the program's unit tests. She found that the students were divided as follows: 40% High, 50% Medium, 10% Low. Since she wanted to give 20 questionnaires, she randomly selected 8 High, 10 Medium, 2 Low students to receive the measure. This meant the sample contained the same proportion of respondents in each category of achievement level as did the total population.

•Matrix Sample. This is a sampling technique in which a portion of the population receives a portion of the measure. This procedure is similar to the others already described in that not all members of the population are tested, interviewed, or otherwise measured. It differs because not all the items in the measure are administered to one individual in the sample. Instead, some members of the population receive some of the items in the measure. The result is that all items are administered to some respondents. How many times each item is given depends on the total number of items, total number of people in the sample, and how many items are being given to each person. Items per person, in turn, may depend on time and number of people included in the sample (with more people less items per person may be necessary).

This technique results in shorter time to complete the measure, a major advantage given the usually crowded school schedule. Another advantage of matrix sampling is that larger item domains can be measured. If the domain of a spelling test, for example, includes 200 words, no other technique would feasibly give information about all the words. But, in a matrix sample, all 200 words can be administered.

Matrix sampling, however, does have some limitations. Since different individuals are receiving different sets of items, comparisons between individuals cannot be made. While this can be an advantage because individuals may not get test anxiety if they know no comparison will be made, it may be a disadvantage because teachers may want to have information on individual students.

For purposes of program evaluation using a large number of items, matrix sampling is highly recommended despite the logistical problems of administering and scoring different forms of a measure.

"WHEN"

In addition to the decision of who will receive the measure, the evaluator and decision makers must be concerned with when, in relation to the program, the measure will be administered. These temporal considerations divide the measures into 3 categories: (1) pre-measures, (2) post measures, and, (3) interim measures.

"WHEN"

Premeasure. A premeasure, given before the program begins, is useful gathering baseline information. That is, what knowledge, skills, or attitudes do individuals have prior to the program? Pretest information will allow a comparison to be made after the program to see if there was any change, positive or negative, over that period of time. A premeasure also ensures that the group does not already possess the terminal skills. If that is the case, a reassessment of objectives is in order.

Postmeasure. As the measure given at the end of the program, a post measure is used to assess program effect. If a premeasure was used, the postmeasure, to allow comparison, should be a parallel form. It is important that the comparison of pre- post program performance is made on the same objectives measured the same way.

A premeasure is not always given. Though less desirable, performance can still be assessed with only the postmeasure.

Interim measures. These are measures given during the program to track enroute skills or attitude changes over the duration of the program. Collecting interim information becomes especially important in formative evaluation, since the evaluator may wish to suggest changes in the ongoing program due to information found from the interim measures.

DESIGN PLANS

Design plans specify the "who" and "when" of the instruments. There are a number of designs commonly used in formative and summative evaluations. Three factors which are manipulated in various designs are:

(1) comparison groups, (2) randomization, (3) pretest.

•Comparison Groups. Many evaluation designs include two groups, the experimental and the comparison group. The experimental group receives the program while the comparison group is given another program or no program. In the summative evaluation of the biology program at Hamilton High, comparisons were made between the four classes receiving the new biology program and two which were receiving the traditional program. At Hamilton High, as in most schools, the comparison group received some alternative program rather than no program. It is not usually possible or advisable to withhold a core program, like biology, math, or reading from students.

It is also appropriate to use comparison groups to answer some of the evaluation questions asked in formative evaluations. For example, to gather information on the effectiveness of alternative sets of supplementary readers in an elementary reading program,Parker Elementary School adopted the following strategy. One group of second grade students who were enrolled in the program received one set of readers while a second group in the program read the other competing set of readers. Performance of the two groups was then compared and the school purchased one of the sets of readers.

•Randomization. Randomization refers to assignment of students to groups. As in a random sample, already discussed, each person has an equal chance of being assigned to either the experimental or a comparison group. Statification, also discussed in sampling procedures, also applies to group assignment. It may be desirable to have the same number of students in the experimental and comparison groups that are equal in reading ability, aptitude, or whatever other factor the evaluator feels might affect performance and should therefore be equally distributed or controlled.

Though randomization is recommended, it is not always possible to implement. Often classes of students must remain intact as a group. If this is the case, the evaluator should try to assess the makeup of the groups involved. If the groups are homogeneous; that is, similar on critical factors (e.g., aptitude) the evaluator may decide that comparisons between groups could be fairly made. The comparison group in this case would be called a non-equivalent comparison group since it was not made up by random assignment. If the groups are not homogeneous, this will effect what can be said about any performance differences that may be found. For example, Edwards was called in to evaluate two English Literature courses in a high school. Unfortunately he was called in 2 weeks after the semester had begin. The students were already assigned (not randomly) to the two courses and no pretests had been given. Edwards, after studying school records, found that students in one class had consistently higher reading scores. Since reading ability could be a significant factor in an English Literature course, Edwards had to tell the decision makers that the evaluation would be severely hindered by this confounding, and that drawing conclusions from any performance differences would be difficult. He suggested delaying the evaluation until the next semester at which time students could be assigned to the 2 courses with attention paid to equalizing reading scores of the groups.

•A pretest. This serves as a check on groups' comparability. If students were not randomly assigned to groups but do receive a pretest, the evaluator will be able to tell if, on the average, the groups are at a comparable level prior to the program. If not, as in Edwards' experience with the English Literature evaluation, the evaluator is limited in generalizing any statements about program effects on performance.

.Formative Considerations for Design

As in the case of evaluation questions and measurement devices, the focus of the evaluation also controls the evaluation design.

Since formative evaluation focuses on a single program, appropriate evaluation designs do not usually include comparison groups. Therefore designs used frequently are:

Name of Design	Control Group?	Random Assignment?	Pretest?
Formative Designs:			
1. Pretest-Posttest	no	no	yes
2. Posttest only	no	no	no
3. Program tracking (case study) (pretest, interim, posttest)	no	no	yes

In any evaluation a combination of these designs may be used depending on the diversity of the measures. For example, a pretest-posttest or posttest only control group design might be used to assess student attainment of major program objectives. These designs would also be appropriate for answering research-type questions. Program tracking designs, however, are most appropriate for measures which are frequently administered throughout a program (e.g., unit tests, observations, interviews, questionnaires). Whichever designs are utilized, administration schedules for the measures may have to remain fairly flexible due to such factors as individual classroom differences. For example, teachers will typically complete units in the program at different times. Therefore, the evaluator's designs for unit measures should be structured so that they can be administered when a teacher informs the evaluator that the unit or instructional sequence is completed.

Summative Considerations for Design

In summative evaluation, often the main concern is how well the program is meeting its objectives and how it compares to other programs. Unlike formative evaluation, a summative evaluation generally has fewer process measures. Since the program should theoretically be running as planned, only a minimum number of interim measures need be used to check on program processes. Some pupil progress measures such as unit tests may be desired to keep tabs on individual progress, but major emphasis will probably be on program comparisons. Therefore, most designs for summative, unlike those for formative, make use of comparison groups. The chart below lists some of the more common designs used in summative evaluation.

Name of Design	Control Group	Random Assignment	Pretest
1. Pretest-posttest control group	yes	yes	yes
2. Posttest only control group	yes	yes	no
3. Pretest-posttest non-equivalent control group	yes	no	yes
4. Time series	no	no	yes

The first three designs on the chart, pretest-posttest, comparison group and non-equivalent comparison group have already been discussed individually. Time series, the last design on the chart, differs from these in the number of times a measure is given. The same measurement is made on the same group of students at regular intervals before and after the program. In this way students serve as their own control. By observing if program X affects the trend in the results, it is possible to determine if program X had an impact. For example, an elementary school attendance counselor wanted some evidence that her small group counseling techniques were effective with students who had a history of unexcused absences. By checking the existing records for each for three months prior to the counseling sessions, she figured that the students were missing an average of six days per month. She compared this with the three months after the counseling treatment and found that the students were absent an average of two days per month. From this and other corroborating information, she was able to show that the counseling technique seemed to have an effect on absences.

EXERCISE 5

Selecting A Design

A meeting of the planning committee has been called to discuss the design or data collection plan of the evaluation measures. Lionel Edwards has brought a chart with the list of evaluation questions and measures already selected by the committee at their previous meeting. Part of this chart is presented on the next page. One question from each category of evaluation questions is listed.

The committee is made up of:

- .Bruno Freitag,Superintendent
- .Evert Walker, Principal
- .Sandi Silverthorn, Curriculum Specialist
- .Lionel Edwards, Evaluator
- .Biology Teacher

Assume the role of one of these committee member (refer to character summaries in Section B). Then, with your team members, discuss and select an appropriate design for each measure. When choosing the design, keep in mind such factors as data collection time, personnel availability, and materials cost.

EXERCISE 5

Selecting A Design

Evaluation Question	Measure	Audience	Who/What (of the audience)	When
1. Outcome: Do students show a positive attitude toward and interest in biology as evidenced by:	Questionnaires Interviews Records	Teachers		
(1) requests to use biology materials (2) voluntarily reads books on biology (3) does biology activities and experiments at home (4) volunteers to bring in materials (5) volunteers to do research	Questionnaires Interviews	Students		
2. Implementation: How much subject matter content is being covered by individual teachers?	Records Observation	Teacher		
3. Cost: What is the cost of wear and tear on laboratory equipment?	Records	District purchising department		
	Questionnaires Interviews	Teachers		
4. Research: Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit is more effective in improving student performance?	Unit retests and Program Posttests	Students		
5. Unanticipated Outcome: Have any logistical problems been caused by the field trips?	Records	District accounting office		
	Questionnaires Interviews	Teachers		

Section G

OUTLINE

EVALUATION DESIGNS

I. Sampling

- A. Random sample
  - .each person in the population has an equal chance of being included in the sample
- B. Stratified random sample
  - .population is divided into catagories and people randomly selected from each catagory in proportion to what is found in population
  - .major advantage: increases likelihood of a representative sample
- C. Matrix sample
  - .a portion of the population receives a portion of the items
  - .advantages: less testing time, larger domains can be measured, limits examinee test anxiety
  - .disadvantages: failure to gather data on individual performance

II. Measures

- A. Premeasure
  - .gathers baseline data
- B. Postmeasure
- C. Interim Measures
  - .track enroute skills, attitude changes over duration of program.

III. Design Plans

A. Comparison Group

.receives alternative program

.appropriate for summative evaluations, and for formative evaluation research questions

B. Randomization

.recommended, not always possible; alternative design is nonequivalent control group

C. Pretest

.assesses comparability of groups

IV. Formative Designs

A. Pretest - Posttest: one group or a sample receives pretest and posttest

B. Posttest Only: one testing given to one group or a sample at the end of treatment

C. Program Tracking: one group or a sample receives pretest, one or more interim tests and posttest.

V. Summative Designs

A. Pretest - Posttest Control Group: students are assigned randomly to groups, both experimental and comparison group take pre- and posttest

B. Posttest - Only Control Group: students assigned randomly to both groups; both experimental and comparison groups take posttest

C. Non-equivalent Control Group: no random assignment, two intact groups are compared, both groups received a pre- and posttest

D. Time series: one group of students who act as their own control, series of measures taken before treatment are compared with a series of measures taken during and/or after treatment.

Analysis. Once the measures and their design are determined, preliminary plans can be made for the analysis. It is recommended, especially in formative evaluations, to "tryout" samples of the analysis explanations and displays with sample members of the audiences. Too complex analyses that are not understood by the audience are not worth the time and effort involved in their preparation. A planned analysis also avoids the problem of having a design that does not lend itself to usual statistical procedures. If this problem is recognized before program implementation, the design and/or analysis can be modified.

Reporting. Planning the report before implementation of the program has even begun may seen premature, but this early preparation can avoid many problems. If, especially in a formative evaluation, the reporting schedule is agreed on before the fact, there will be no misunderstandings about when and how many reports are expected. The audiences for the reports and the content for each audience can also be resolved in this planning stage. A number of different people (e.g. administrators, teachers) can be involved in these decisions and feel that their input is valuable. This involvement could result in a more positive attitude or a stronger committment toward the new program.

SECTION H

Section H

DATA COLLECTION

PLANNING THE EVALUATION
DOCUMENTING THE EVALUATION

This section presents guidelines for data collection, describes some common problems which arise during data collecting and suggests possible strategies for avoiding or coping with these problems.

Guidelines for Data Collection

.Orientation

.keep groups affected by data collection informed about, and if possible involved in, decisions regarding data collection

.inform groups about the purposes and requirements of the evaluation

.agree on the data collection schedule and procedure

.agree on the format and deadlines for reporting

.make data collection as unobstrusive, economic and convenient as possible

.During data collection

.anticipate and hopefully avoid problems during data collection by keeping in touch with respondents

.remind respondents about the data collection schedule

.check that scheduled data collection is synchronized with program activities

.be unobstrusive. Try to disrupt normal classroom activities as little as possible.

Common Problems in Data Collection

Although following the guidelines included in this section can help avoid some of the problems that arise as the evaluation is conducted, problems relating to instruments, design, analyses and reporting do occur

.Instrumentation Problems

.poorly constructed tests directions may be confusing, or items may be at the wrong difficulty level. Both problems can be avoided by having tests reviewed by teachers.

.instruments are not administered in a standardized manner. Teachers may give students helpful hints in completing the test.

Design Problems

.Sampling. Sometimes the sampling plan is not implemented due to logistical problems (e.g., not enough students of each required ability level enrolled in the program) or confusion about its importance, desirability or feasibility.

Avoid problems by conferring with personnel on design during planning, or if too late, sort out in analyses.

.Scheduling. Schedule of measure administration may be disrupted by logistical problems (e.g., materials not arriving on time, uneven pacing, disruptive school activities). Careful documentation of program may determine if test data is useable.

Analysis Problems

.sometimes intended audiences cannot understand the data as presented. Keep simple and set up discussion meetings.

132

.sometimes the data are not timely. Try to agree on a reporting schedule before the evaluation. Keep data simple so analyses does not hold up the report.

Problems with Reporting

.Recurring complaints about reports include: program staff do not attend reporting session; no one reads, understands or agrees with reports; the reports are not on schedule. Try to avoid these problems by involving program staff in negotiating format and content. Set up meetings to discuss completed reports.

SECTION I

Section I

Data Analysis - Summarizing Evaluation Data

PLANNING THE EVALUATION
CONDUCTING THE EVALUATION
DOCUMENTING THE EVALUATION

I. Data Analysis - Summarizing Evaluation Data

- Descriptive Statistics
- Tests of Differences
- Tests of Relationships
- Formative/Summative Considerations

J. Reporting

ANALYSIS--SUMMARIZING EVALUATION DATA

The last section of the evaluation plan should outline the planned analysis of the data collected from each of the measures. To simplify, the kinds of analysis commonly used in evaluation will be presented in three categories: (1) descriptive statistics (2) tests of differences (3) tests of relationships.

Descriptive Statistics

As the name implies, these statistics describe the information or data collected. Often an evaluator is faced with many individual scores, (e.g., total items correct on a test, total or average ratings on a questionnaire, tallies on an observation sheet). Looking at any of these numbers, it is difficult to interpret the information contained in them. Reducing the numbers into manageable and meaningful statistics allows the evaluator and others to understand just how the respondents performed on the measure. For example, the posttest scores for one class in the biology program are ordered below:

Biology Posttest Scores for Classroom #3

99	91	84
99	89	84
98	89	82
98	87	79
96	87	77
95	86	77
94	86	74
92	86	68
92	84	62
92	84	30

N=30
Range=37
Mean ($\bar{X}$)=86.6
Median=87
Mode=84
Standard deviation ($\sqrt{s}$)=9.04

The first thing Edwards did with these scores is to look for outliers--scores that are clearly outside the pattern. In this case, the lowest score, 30, is 32 points below the second score. Since this seemed peculiar, Edwards checked with the teacher and found that this student had been absent many weeks during the semester and thereby missed much instruction. Edwards decided to eliminate this student from the analysis. Now he is ready to compute some simple statistics. A number of statistics are easily derived from these raw scores. Some of these include: range, average, and standard deviation.

The range is the difference between the lowest and highest scores.

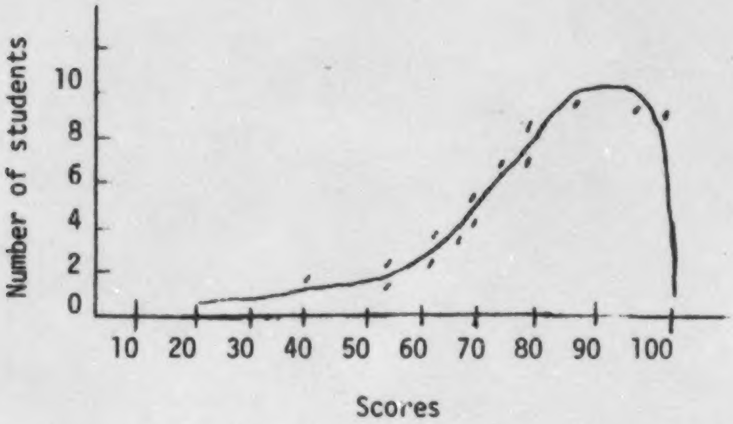
The average in statistical terms may be the *mean* -- the statistically average score, the *mode* -- the most frequent score, or the *median* -- the score above and below which 50% of the students score.

The standard deviation, another useful statistic, shows how much the scores are spread out around the mean or how much variability is in the scores.

Once Edwards has this information, he must decide the best way to present it to the various decision audiences. Descriptive statistics are easily amenable to pictorial representation in the form of tables, charts and graphs. These provide an effective visual image of the data.

For example, the chart below was one of several shown to teachers at one of the monthly meetings during the formative evaluation. The chart summarizes one classroom's scores on a unit posttest. Since the program is only being used in four classrooms, Edwards can take the luxury of giving this detailed information. But he would probably also present a graph combining the same data for all four classes. In larger scale evaluations, summary data on all classes is often the only feasible way to present the information.

Class 1: Posttest Scores on Unit 3



Tests of Differences

If the evaluation wishes to compare the performance of different groups, either within a program as in formative evaluations, or between programs, as in summative evaluations, the usual analyses procedure is a test of differences. There are a variety of difference tests, the appropriate one depending on such factors as randomization, sample size, and type of measurement data. However, all difference tests measure the amount and importance of the difference between the means of two sets of scores.

Since identical average scores between groups would be very unusual, there is always some difference to analyze. A test of differences examines the mean difference for significance. A statistically significant difference means that the difference between the means is not due to chance but to some other factor, presumably the program or a facet of the program under evaluation. As an example of one test of differences, let's look at one used by Lionel Edwards in the formative evaluation to study the effects of number of field trips on achievement. He compared the laboratory posttests of the students who took two field trips to those who had taken one. The following chart summarizes this information.

Comparison of Achievement by number of Field Trips

	N	2 field trips		N	1 field trip		T-test
		$\overline{X}$	S.D.		$\overline{X}$	S.D.	
Posttest	60	78.6	11	60	78.1	10.8	NS*
Posttest	60	89.4	12	60	88.3	11.3	

*NS = Not significant

Using a test of differences (a t-test in this case), Edwards found that the number of field trips did not seem to make a significant difference in posttest scores. This information will be in the final report and help decision makers decide how many field trips to include in the program in the future.

Tests of Relationships

This third category includes those statistics which look at the relationship between two variables. That is, what is the likelihood that if X varies, Y will vary in a systematic way. For example, one question Edwards wishes to answer is: "Was student achievement in the biology course related to student attitude?" Edwards wondered if attitude toward the program would be more positive the higher the student's performance on the program posttest. To find out if this hypothesis was true, he listed, in rank order, each student's score on the posttest and each student's total rating of the program from the post program attitude questionnaire. For brevity, scores from only one class are shown below. By looking at the scores Edwards could see a tendency for questionnaire ratings to increase as test scores increased.

Figure : Relationship of Achievement in and Attitude Toward Biology

Posttest Score	Questionnaire Rating
54	11
55	10
61	11
68	14
72	13
74	12
74	15
77	14
78	12
81	21
89	17
90	12
91	18
91	20
92	19
94	15
94	18
94	24
96	23
98	22

N=20
Possible posttest score=100
Possible questionnaire rating=25
r=.774

Edwards then computed a correlation, the most common of the tests of relationships, and found a correlation (r) of .774. (The highest correlation possible is 1.00). Edwards is now able to say that there seems to be a tendency that as student's performance on the posttest increased a positive attitude toward the program increased.

Note that Edwards only said that his statistic has shown a relationship between the two factors, achievement and attitude. This is not to say that one is the cause of the other. The relationship shown between achievement on the biology test and attitude may be reflecting a relationship between achievement and attitude toward school that existed before this new program. So both factors could be due to or caused by a third factor. A correlation does not prove causality. For example, at one time farmers noticed that the crops were usually ripe the same time geese were flying overhead on their journey south. The more geese, the more likely that the crops were ripe. Therefore, there was a high correlation between the number of birds flying overhead and the acreage that was ready to be harvested. However, this does not mean the birds caused the crops to ripen, nor does it imply that ripe crops caused the birds to fly south. Instead, a third factor accounts for both--autumn. The closer to autumn, the more geese flying south, and the more crops that are ready for harvest.

Tests of relationships then, are useful in evaluation to show how two factors relate to each other; to see if there is a tendency for two factors to vary together in a systematic way.

Formative considerations for Analysis

One of the major purposes of formative evaluation is to provide feedback to the people running the program (e.g., teachers), and people making decisions about the program (e.g., administrators, school board members). Therefore, the evaluator must be concerned with analyzing his data in a way that is simple, easy to understand, and useful. This is especially true for informal interim reports or meetings. Descriptive statistics such as means and percentages are familiar to most people and will probably be most favorably received. There is no reason to overwhelm an audience with sophisticated analyses that are not essential. In fact, the unwarranted use of sophisticated analyses may antagonize and alienate people.

The method of data presentation is as important as the type of analyses done. The evaluation should present teachers and administrators with easy to read summary data about how the program is running and how students are performing. Graphs, tables and other visual displays of descriptive statistics are highly recommended for these interim reports. The data presentations should include not only test results but summaries of other measures taken - questionnaires, observations, etc.

The level of detail used to report information will probably depend on the program. The evaluator may wish to show student performance on specific objectives within units rather than just whole unit test results. The frequency of these reports will be discussed in the reporting section of this workshop.

In the final report for a formative evaluation, the evaluator will want to expand the analysis strategy. Descriptive statistics, should be included though probably on major program outcomes rather than small sequences of instruction. In addition, the final report will include results of any pilot tests conducted to look at different components of the program. For these results, tests of differences and tests of relationships will also be appropriate to report.

Summative considerations for Analysis

Unlike the analysis for a formative evaluation, the summative analysis will include less process information. Since the evaluator will not be giving feedback to change the on-going program, but will be summarizing the established program, only major activities need be documented.

The analysis for a summative evaluation may focus on comparative information concerning the target program and any program or programs to which it is being compared. This information should include not only descriptive statistics but also test of differences and tests of relationships whenever possible.

Exercise 6

Data Analyses and Interpretation

Four evaluation questions from the Hamilton High Biology Program are listed on the following pages. The data Edwards collected is presented below each question. Working in teams you will be asked to decide how to display the data, interpret the data already given, or make recommendations based on the data. Directions are given for each question.

1. Outcome Question #1: Do students show a positive attitude toward and interest in biology as evidence by:

   (1) requests to use biology materials
   (2) voluntarily reads books on biology
   (3) does biology activities and experiments at home
   (4) volunteers to bring in materials
   (5) volunteers to do research

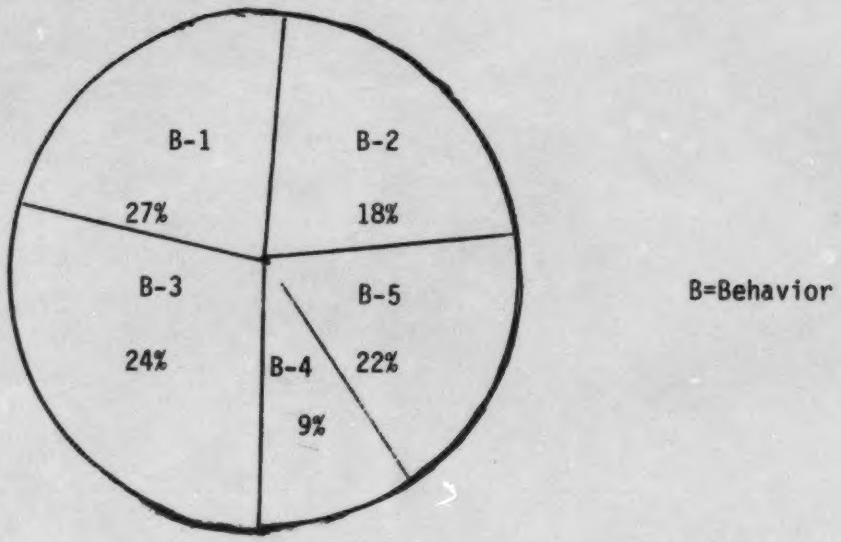
Using the information from the teacher records, Edwards compiled a table. Questionnaires and interviews substantiated the accuracy of the record keeping. The table shows how many times each behavior was exhibited by each student. A portion of that table is presented below.

	Activity				
Student	1	2	3	4	5
1	//	/		/	
2	~~////~~	/		//	
.					
.					
.					
.					
.					
120	/	//		/	////

Next Edwards must put this data from the table into a more meaningful form so that it can be presented to and understood by the people reading his report.

Working with your team members discuss and then list (in the space below) suggestions for organizing and displaying the information in the table. If time permits illustrate with pictorial displays.

Example 1. Calculate the percentage of students displaying each behavior and display in a pie.



2. Cost Question: What is the cost of wear and tear on laboratory equipment?

The following table presents the information Edwards collected from the records of the district purchasing department and from teacher responses on a questionnaire and interview.

Nature of Expenditure	Dollar Expenditures	
	Spring Semester Last Year (old program)	Fall Semester This Year (new program)
Equipment Replacement*	$425	$350
Equipment Purchases**	$225	$400

---

*Replacements include slides, new microscope (to replace broken), specimens for dissection

**New Purchases include lab manuals, slides, new microscope, glass equipment.

How would you interpret this information? What explanation could you give for the costs in each block of the table? What conclusions, recommendations would you make about the laboratory equipment?

3. Research Question: Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit, is more effective in improving student performance?

To answer this question in his final report, Edwards looked at two factors. First, he compared the program posttest scores of students using the Practice Kit and students using the Instructional Kit. The following table presents the results.

Program Posttest Scores of Students
Receiving Supplementary Practice

Kit	N	Posttest Score		
		$\overline{X}$	SD	T
Practice Kit	24	72	2.9	t=3.7*
Instructional Kit	22	75	5.7	

$\overline{X}$ = mean score
SD = standard deviation
* = significant

A t-test revealed that there was a significant posttest performance between groups.

Edwards also included the monetary cost per kit.

Practice Kit 3.00 per pupil
Instructional Kit 4.50 per pupil

The cost information had already been included in the cost section, but Edwards felt it was an important factor when considering the cost-effectiveness of the kits.

In the space below, write an interpretation of the table and make recommendations on the use of the kits for next semester.

4. Unanticipated Outcome: Have any logistical problems been caused by the field trips?

To help answer this question all tenth grade teachers were asked to keep records of when students from their class were absent due to biology held trips as opposed to other field trips.

The data from the questionnaire given to all tenth grade teachers is presented in the table below.

TABLE

Logistical Problems Due to Field Trips
(Portion of post program questionnaire)

Question	Number of Teachers (N)	Response Frequency (%)
C. Did many students miss your classes as a result of the biology field trips?	15 (all sophmore teachers)	
1. yes		65%
2. no		35%
D. How do you react to students missing your classes because of the biology field trips?	15	
1. favorably		28%
2. unfavorably		72%
E. Did you find you had to do extra work to prepare for those students who missed your classes because of the field trips?	15	
1. quite a bit, somewhat		80%
2. not at all		20%
F. Were all students able to participate in the field trips?	4 (biology teachers)	
1. yes		0
2. no		100%
G. If not all students were able to participate in field trips, why were they unable to participate?	4	
1. absences		11.8%
2. conflicts with other classes or school activities		81.2%
3. students unwilling to go		7.0%

District accounting office records were also examined. Edwards found that buses did not arrive on time 5 times, even though the bus company had a schedule of all times buses were needed; three buses broke down returning from trips; and one trip had to be postponed due to unavailability of any bus.

What conclusions would you, as the planning staff, draw from this information? Is there any information you would wish to have to aid your recommendations about field trips?

Additional Information Desired:

Conclusions:

Recommendations:

SECTION J

Section J

Reporting Evaluation Results

PLANNING THE EVALUATION
CONDUCTING THE EVALUATION
DOCUMENTING THE EVALUATION
  I. Data Analysis
  J. Reporting Evaluation Results

Strategies used in writing evaluation reports must be appropriate for the various decision audiences who will read them, especially those sections containing discussions of design and analysis.

The style and length of the report should also vary, again depending on the audience. For example, a detailed technical report would be presented to a funding agency, school board, or teachers whereas a less technical leaflet summary highlighting program results might be prepared for presentation to parents. Styling the content and format of the report for each decision audience will help ensure that the report will be read and have maximum influence.

The following outline lists the points that should be included in an evaluation report. Since each evaluation differs in content and purpose, every point may not necessarily be included in the report, conversely, there may be points not mentioned in the outline, that seem appropriate to include. The outline is presented as a guide which should help an evaluator avoid overlooking information audiences may need to fully understand the results of the program. A clear report will also increase the audience's confidence in the conclusions and recommendations.

Outline of An Evaluation Report

	Suggested organization of the report	Suggested order in which the sections should be written
Section I.	Summary	last
Section II.	Purposes of the Evaluation	first
Section III.	Background Information Concerning the Program	second
Section IV.	Description of the Evaluation Study	third
Section V.	Results	fourth
Section VI.	Discussion of the Program and Its Results	fifth
Section VII.	Conclusions and Recommendations	sixth

.Section I. Summary

The summary is a brief overview of the evaluation report designed for the person too busy to read the entire report. It is usually no longer than one or two pages. It should explain why the evaluation was conducted and include a list of majors conclusions and recommendations. Although the summary is the first section of the report, it is the last section that should be written.

.Section II. Purposes of the Evaluation

This section describes why the evaluation was conducted, what it was intended and what it was not intended to accomplish. Section II describes the assignment that the evaluator accepted and for this reason it is recommended that the evaluator prepare this section early in the negotiations for the evaluation. A draft of this statement or letter of agreement should be agreed upon by all interested parties and be kept on file.

Section III. Background Information Concerning the Program

This section sets the program in context. It describes how the program was initiated and what it was supposed to do. If the audience for the report has no knowledge of the program, the program should be described in detail. If the report is to be read by those familiar with the program, section III may be brief, stating the facts "for the record."

Drafting section III at the time the evaluation is planned will ensure that the evaluator has a clear concept of the program, and what it is intended to accomplish. This draft may then be circulated to program personnel for their additional comments.

The typical content for this section includes:

A. Origin of the program (how and why the program was started)
B. Goals of the program (why and how goals were established)
C. Characteristics of the program (materials and activities)
D. Students involved in the program (why and how selected)
E. Faculty and others in the program

Section IV. Description of the Evaluation Study

Included in this section is a description of the methodology of the evaluation, including the design for each evaluation question. It is important that these descriptions be detailed. If people are to have faith in the conclusions of the evaluation they need to know how information was obtained.

Samples of all instruments should be included (with the exception of commonly published tests or tests which are not to be reproduced).

As indicated in section III, section IV should be drafted when the evaluation is being planned. It is advisable to circulate a copy of this draft to program personnel. Getting people to agree on the evaluation plan will increase the credibility of the results.

.Section V. Results

This section presents the results of the various measurements which were to (a) document what happened, (b) measure the outcomes and state how the program approached its goals. If the measurements were reliable and valid, these results are the "hard data." Anecdotal evidence or testimonials are the "soft data" which are also to be included in this section.

The results section should not be written until all the data has been analyzed, recorded in tables, graphed or plotted, and tested for significince where appropriate. When all the data has been summarized and readied for display, the evaluator may then proceed to write section V, presenting and explaining each table and display of data.

.Section VI. Discussion of Results

This section of the report could be included in section V. However, if the program and/or the evaluation is very complex it may be advisable to make the discussion and interpretation of results a separate section of the report. The results should be discussed with particular reference to section II, Purposes of the Evaluation.

The typical content of this section includes discussions of the following questions:

- .How good were the results of the program?
  - .Comparison of program results with what might have been expected had there been no program
  - .If a control group was used, were program results better than control group results? Was the size of the difference statistically significant?
- .How certain is it that the program caused the results?
  - .Other possible explanations.
  - .What confounds need be considered?
- .What are the dollar and non-dollar costs of the program?
  - .What extra monies were required to implement the program? How might monies have been used for if not for the program?
  - .What are the costs of human resources; loss of alternative opportunities?
- .What are the benefits associated with the program-dollar and non-dollar?

When appropriate, a "Cost-Benefit Summary Table" in which costs and benefits, both dollar and non-dollar, should be presented at the beginning or end of this section. This will enable the decision maker to decide to what extent incurred costs justify derived benefits. Operating Costs should be distinguished from start up costs since the latter will not be needed if the program is continued. Also, costs and benefits for which the evaluation has shown sound evidence should be distinguished from suspected costs or benefits not substantiated by objective procedures.

Section VII. Conclusions and Recommendations

Section VII is the most influential part of the evaluation report. This last section is best presented in the form of a list rather than in narrative form. Be sure to emphasize what conclusions are to be "tentatively" or "firmly" held. Care should be taken to attend to all concerns raised in section II which described in detail the purpose of the evaluation.

Formative Considerations in Reporting

In a formative evaluation, the evaluator will need to prepare several informal reports in addition to the final report due upon the completion of the program. Although he may choose to vary the style of the informal reports (e.g., oral presentations to the entire program staff, written briefs) their prime function is to supply frequent feedback to the program staff about how the program is shaping up. These reports are diagnostic and trouble shooting in nature since their purpose is to assist in program improvement. The number of reports will undoubtedly be determined as much by time, money, and personnel availability as evaluator initiative.

It is recommended the evaluator furnish a report upon the completion of each program objective or major instructional unit in order to accurately inform the staff concerning the degree to which program processes are helping achieve program goals. This immediate feedback facilitates program improvement by allowing faulty processes to be modified quickly. After completion of the program, the evaluator would then submit a formal written report.

These reports should present a detailed description of the program thus far, and then, based on the findings, offer some recommendations for improvement. When possible, these recommentations should reflect the concerns of the program staff. The recommendations can be arrived at with the help of the staff or at least reviewed by them. The report should specify the extent to which objectives are being met, by which "types" of classes and at what point in time within the anticipated program schedule, and how procedures are being implemented and the effect these procedures are having. If original designs have been implemented differently the report should state what adjustments were made and how they affected the measurement and analysis considerations.

Summative Considerations in Reporting

The reporting strategy used in summative evaluations is more formal than that employed in formative evaluations. Typically the evaluator prepares a written report at the end of the program although one or more formal interim reports are sometimes submitted. Another major difference between the reports prepared for formative and summative evaluations is the reporting audience. Typically summative reports are intended for use by a funding agency which must decide whether or not to refund a program. Therefore, summative reports are more politically loaded than formative reports since an unfavorable report may result in a decrease or elimination of funding for a particular program.

Since evaluation, both formative and summative, is designed to provide information for program improvement, a summative evaluator may wish to include suggestions for future modifications in the program. However, due to the political sensitivity of summative reports, the evaluator must adapt strategies which allow him to make recommendations which have a likelihood of being implemented. The evaluator might present recommendations to the program staff for their review before submitting the recommendations to a funding agency or the recommendation might only be included in a seperate internal report for the program staff's use.

Final Formative Report

This section presents an example of the type of language and information included in a final report of a formative evaluation. The report has been condensed and will discuss only one evaluation question within each of the five catagories: program outcome, implementation, research, cost, unanticipated outcome.

I. SUMMARY (to be completed in exercise 7)

II. PURPOSE

This report of the formative evaluation for Hamilton Unified School District's biology program contains information about the biology program implemented at the beginning of the school year 19__. The report documents this evaluation so that decisions can be made to modify the program. The following list of questions formed the basis of this evaluation.

Outcome

1. Do students show a positive attitude toward and interest in biology?
2. Are students able to explain various interactions between organisms and their environment?
3. Are students able to explain or demonstrate various chemical processes necessary for life?
4. Are students able to classify organisms according to phyla and list the characteristics of various phyla?
5. Are students able to describe the structure of an organism and state how its structure relates to its function?
6. Are students able to list strategies to minimize environmental pollution?

Implementation

1. Are the field trips being conducted as planned?
2. How much subject matter content is being covered by individual teachers?
3. Are teachers following the program's planned instructional sequence?

Cost

1. Is the performance of students using the Practice Kit or the Instruction Kit significantly enhanced relative to the cost of the kit?
2. What is the cost-effectiveness of field trips?
3. What is the cost of wear and tear on laboratory equipment?

Research

1. Which remedial technique included in the Living Systems program, the Practice Kit or the Instruction Kit, is more effective in improving student performance?
2. What is the effect of the frequency of field trips on student attitude?

Unanticipated Outcome

1. Have any logistical problems been caused by the field trip?
2. What other outcomes can be attributed to the biology program?

III. BACKGROUND INFORMATION:

A. Origin. As a result of concern on the part of parents, students and teachers, Hamilton Unified School District realized the need to improve student performance, attendance and attitudes in and toward school by making classes more relevant and practical in terms of students' immediate and future needs. After the general needs assessment of the district, which revealed five major academic deficiencies, members of the school board decided to focus their attention on the revision of the tenth-grade biology program at Hamilton High. A committee was formed to select a program. The committee included teachers, administrative personnel, and community representatives. It was decided that a formative evaluation would be conducted during the program's first semester of operation.

B. Goals. The major goals of the biology program, Living Systems, are:

- .To improve student attitude toward and interest in biology
- .To improve students' understanding of the relationship between organisms and their environment
- .To improve students' understanding of the chemical bases of life
- .To improve students' understanding of reproduction, heredity, and genetics
- .To improve students' understanding of major classification of organisms
- .To improve students' understanding of the relation between the structure and function of organi
- .To improve students' understanding of man's responsibility contribution to the environment.

C. Program Objectives are:

- Students will display a positive interest in and attitude toward biology by any of the following behaviors
  - (1) requests to use biology materials
  - (2) voluntarily reads books on biology
  - (3) does biology activities and experiments at home
  - (4) volunteers to bring in materials
  - (5) volunteers to do research
- Students will demonstrate their understanding of the relationship between organisms and their environment by explaining various interactions between organisms in a given environment (e.g., food chain, population dynamics and changes, food production.)
- When presented with a chemical process necessary for life students will explain and/or demonstrate this process (e.g., enzyme active, ATP cycle, termentation, respiration, photosynthesis)
- Students will correctly describe specific processes involved in reproduction and tell how they contribute to the continuity of life (e.g., mitosis, meiosis, DNA replication; Mendelian heredity; natural selection; asexual reproduction,)
- Presented with a drawing of an organism representing a certain phylum, students will identify the phylum to which it belongs and list the characteristics of the organism that are typical
- Presented with an organ or organ system students will describe the structure of the organ or organ system and relate it to the functions it performs (e.g., blood vessels, muscles, fruits and seeds, small intestines)
- When presented with sources of pollution in the local environment, students will list strategies which could result in decreased pollution.

D. Characteristics. The biology program was a one semester course with each class meeting once a day for 50 minutes, five days a week. In addition to the regular class meetings, each student attended a 3 hour lab session once a week. A total of 120 students enrolled in 4 biology classes (30 students in each class) and 4 class teachers participated in the new program. Each of the four classes had an aide. Field trips took place on a staggered schedule. Two classes went on monthly field trips while two went on field trips twice a month. Approximately half of those students in each class who required remedial instruction to attain a passing grade in biology received the Practice Kit while the remaining students received remedial practice through the Instruction Kit.

INSTRUCTIONAL AND EVALUATION MATERIALS INCLUDE:

1. Equipment: microscopes, dissecting kits, specimens
2. Teacher's manual
3. Laboratory manuals
4. Testing materials: ditto masters for chapter pre and post quizzes; unit tests; pretest and final posttest.
5. High school biology textbook entitled "Living Systems". . . .

INSTRUCTIONAL STRATEGIES:

Lecture, class discussion, labs, field trips.

STUDENTS:

The course was designed as a 10th grade introductory biology course.

FACULTY:

The four teachers participating in the new biology program were Paul Schmidt, John Lanich, Maria Roberts and Stephaine Muller. All of them hold Bachelor of Science degrees.

IV. DESCRIPTION OF THE EVALUATION STUDY

Outcomes:

Unit tests and an end of program test which were included in the program were administered to answer achievement outcome questions. Within program effects were determined by unit tests as well as an overall posttest. Pre and post tests for the laboratory sessions were also administered affective measures included questionnaires, interviews, and teacher records.

Implementation:

An observation schedule listing all major points concerning use of materials was used and compared to prepared teacher lesson plans and the course outline developed by the program planners to find out any discrepancies. Interviews were also conducted with teachers and students to find explanations for any discrepancies. The following memo is an example of a discrepancy in implementation that was noticed by the evaluator.

From: Lionel Edwards
To: Biology Teachers
Re: Biology Program Implementation

The biology program appears to be progressing well. However, classroom observations have revealed that in some of the classes teachers are omitting exercises and readings from some of the units. For an effective and accurate evaluation, it is crucial that all the topics or instruction are covered as we agreed in our planning sessions. If any teacher omits topics or changes strategies it will seriously jeopardize our evaluation results. If there are any problems of which I am unaware please contact me as soon as possible.

Of course, if you have any questions or suggestions I'm always at your disposal.

Lionel

Cost:

An assessment of quantative and non-quantitive costs was made through questionnaires and interviews with program administrators and teachers and a review of school accounting records.

Research:

The research question addressing the effect of the Practice Kit vs. the Instruction Kit on student achievement was assessed by comparing the retest performance of those students receiving the Practice Kits to the retest performance of students receiving the Instruction Kit.

Unanticipated Outcomes:

Questions relating to unanticipated outcomes were asked during teacher and student interviews and included in questionnaires to students, teachers, and parents.

V. RESULTS [for this sample report, only one evaluation question is presented for each category]

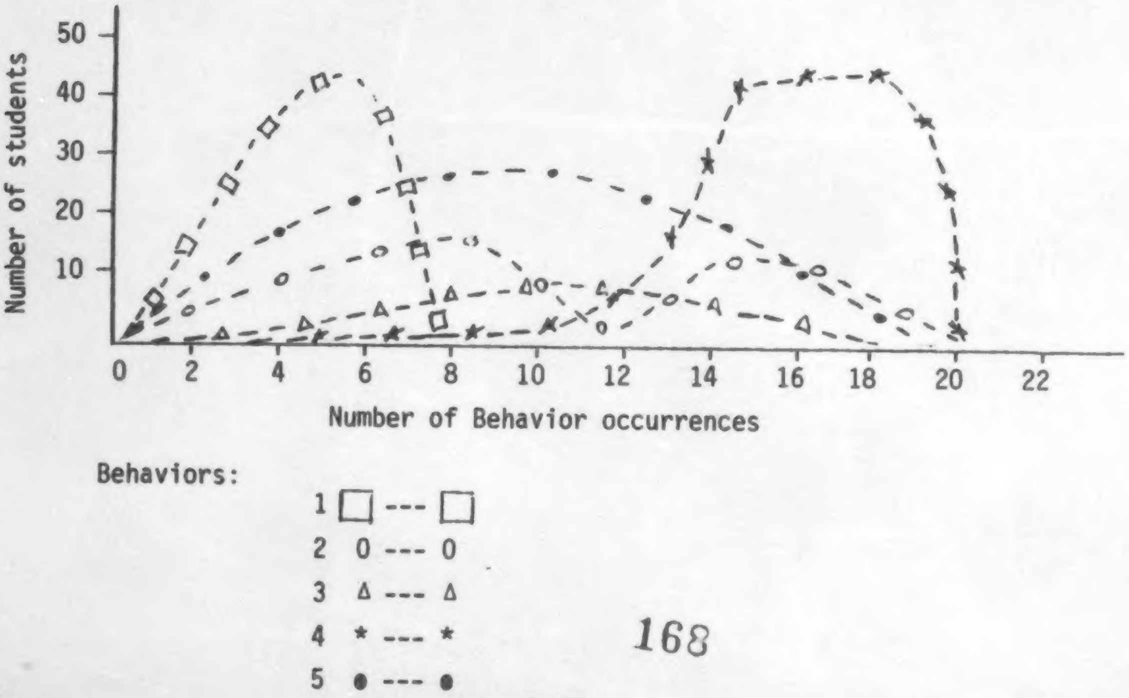
A. Program Outcomes

1. Outcome: Do students show a positive attitude toward and interest in biology as evidence by:

   (1) requests to use biology materials
   (2) voluntarily reads books on biology
   (3) does biology activities and experiments at home
   (4) volunteers to bring in materials
   (5) volunteers to do research

This program outcome was measured by administering questionnaires to students and teachers, interviewing students and teachers, and examining teacher's records. During two interim interviews and on a post program questionnaire all the teachers in the program were asked to characterize students' attitude toward the program and to list incidents with individual students or with the class as a whole which they felt evidenced negative or positive attitudes toward the Living Systems program. In addition, teachers were asked to keep a record of the number of students who engaged in the 5 activities listed in this evaluation question. Records were examined monthly.

The following graph shows the number of occurences of each behavior:



Interviews revealed that teachers felt students in their classes felt positively toward the program and seemed enthusiastic about biology.

At the end of the program a random sample of students (N=40) filled out an anonymous questionnaire asking them how they felt about the Living Systems program. They were also asked to indicate if they had engaged in any of the 5 activities listed in the evaluation question.

Results of the questionnaire substantiated the accuracy of the teacher records.

B. Program Implementation

How much subject matter content is being covered by individual teachers?

This question was measured by observing each teacher's class four times during the program and by examining the records teachers were requested to keep. Records included amount of material covered and any reasons material was not covered. At the completion of each observation, teachers' records of material covered for the preceding time period (approx. 1 month) were examined. Observations were designed to assess the degree to which teachers were following the instructional plan.

The following table shows the percentage of material covered and the amount and reasons for material not covered:

	Teachers			
	1	2	3	4
Lesson done as scheduled	90%	87%	92%	86%
Lesson changed	10%	13%	8%	14%
*Activity omitted	3%	6%	4%	3%
Activity sequence changed	5%	4%	1%	6%
**Activities added	2%	3%	3%	5%

---

*omitted activities were -- student reports
discussion of homework
exercise

**activities added were -- speakers
in-class exercise
filmstrip

C. Cost Question

What is the cost of wear and tear on laboratory equipment?

This question was answered by examining the district purchasing department's records. Lab teachers were interviewed and completed a questionnaire which contained items dealing with student wear and tear of lab materials. Table ___ summarizes the information obtained from the purchasing department's records, comparing the costs of replacing and purchasing equipment for fall semester this year and spring semester last year.

Nature of Expenditure	Dollar Expenditures	
	Fall Semester This Year	Spring Semester Last Year
Equipment Replacement*	$350	$425
Equipment Purchases**	$400	$225

---

*Replacements include slides, new microscope, specimen for dissection

**New Purchases include lab manuals, slides, new microscope, glass equipment.

The results of teacher questionnaires and interviews showed that all lab teachers felt students were being more careful with laboratory equipment this year. This is also reflected in the lower expenditure for equipment replacement this year relative to last year.

Teachers felt that, during its first year at operation, the new biology program is more expensive than the old biology program. However, most felt that these additional expenses are associated with the installation of the new program and will not continue during the next school year. Purchasing department records show greater outlays for additional equipment to year up for the new program.

D. Research Question

Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit, is more effective in improving student performance?

To answer this question, a comparison was made between the program posttest scores of students using the Practice Kit and students using the Practice Kit and students using the Instructional Kit. The following table presents the results:

Program Posttest Scores of Students
Receiving Supplementary Practice

Kit	N	Posttest Score		
		$\overline{X}$	SD	T
Practice Kit	29	72	2.9	t=3.7*
Instructional Kit	22	75	5.7	

$\overline{X}$ = mean score
SD = standard deviation
* = significant

A t-test revealed that there was a significant posttest performance between groups.

E. Unanticipated Outcome Question

"Have any logistical problems been caused by the field trip?"

This question was researched using District accounting records, questionnaires to all 10th grade teachers and interviews to a 25% random sample of teachers of subjects other than biology at Hamilton High. Table __ presents the results of the responses to questions about logistics problems caused by the field trips.

TABLE

Logistical Problems Due To Field Trips

Question	Number of Teachers (N)	Response Frequency (%)
C. Did many students miss your classes as a result of the biology field trips?	15 (all sophmore teachers)	
1. yes		65%
2. no		35%
D. How do you react to students missing your classes because of the biology field trips?	15	
1. favorably		28%
2. unfavorably		72%
E. Did you find you had to do extra work to prepare for those students who missed your classes because of the field trips?	15	
1. quite a bit, somewhat		80%
2. not at all		20%
F. Were all students able to participate in the field trips?	4 (biology teachers)	
1. yes		0
2. no		100%
G. If all students were able to participate in field trips, why were they unable to participate?	4	
1. absences		11.8%
2. conflicts with other classes or school activities		81.2%
3. students unwilling to go		7.0%

Most teachers found that many students were missing their classes because of the biology field trips. This absence meant that the teachers had to do extra preparation to help students catch up on what they had missed. The majority of teachers sampled were generally not pleased with this situation.

There were also problems in arranging for bus transportation due to conflicts in availability of busses as evidenced by the accounting office records. Conflicts in class schedules and school activities played a major role in preventing some students from participating in the field trips.

EXERCISE 7

Formative Report Summary

Based on the evaluation report, list points to include in a summary of the evaluation that will be appropriate for parents. The summary should include a description of:

the program and its goals

the evaluation purpose and method

the evaluation results and recommendations

Summary

SECTION K

EXERCISE I. FEEDBACK

Categorizing Evaluation Activities

Summative Evaluation — 1. The program is well established and will be compared to other programs.

Needs Assessment — 2. The district will poll the community and school personnel for opinions on high priority needs.

Program Planning — 3. A program is being planned.

Formative Evaluation — 4. Data on an innovative program is being gathered to see if modifications should be recommended.

FEEDBACK

EXERCISE 4

Measurement Devices

Date

Dear Susan:

I am sending you a list of the evaluation questions and the measures I will be using to answer these questions. Also included is a rational for my decision. What do you think? Any suggestions or comments would be appreciated.

Evaluation Question	Audience	Measure	Rationale
1. Outcome: Do students show a positive attitude toward and interest in biology as evidence by: (1) requests to use biology materials (2) voluntarily reads books on biology (3) does biology activities and experiments at home (4) volunteers to bring in materials (5) volunteers to do research	Teachers	Questionnaires Interviews Records	Attitudes and interests are best measured by combining both written responses (Questionnaire) with oral responses (Interviews) and examination of behaviors (Teacher Records).
	Students	Questionnaires Interviews	
2. Implementation: How much subject matter content is being covered by individual teachers?	Teachers	Records	Teachers can be asked to keep records of what material is covered each class.
		Observation	Observations of classes will undoubtly be conducted and this can be one item on the observation checklist.
3. Cost: What is the cost of wear and tear on laboratory equipment?	District purchasing department	Records	Records will show monetary costs of repairing and buying equipment.
	Lab Teachers	Questionnaires Interviews	Lab Teachers may give worthwhile information about equipment use not found in the records.
4. Research: Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit.is more effective in improving student performance?	Students	Unit retests and Program Posttests	Comparison of performance on achievement tests should show effectiveness of the kits.
5. Unanticipated Outcome: Have any logistical problems been caused by the field trips?	District accounting office	Records	Records would reveal problems, in cost, scheduling, availability of transportation, etc....
	Teachers	Questionnaires Interviews	Use of a questionnaire would allow for responses, from a large group of people (e.g. many teachers in the school) and save time. Interviews, while time consuming, would provide information not tapped by questionnaires.

Feedback

Exercise 5

Selecting A Design

Evaluation Question	Measure	Audience	Who/What (of the audience)	When
1. Outcome: Do students show a positive attitude toward and interest in biology as evidenced by:	Questionnaires Interviews Records	Teachers	all in program (4 teachers)	Questionnaire - 1 pre-program 1 post-program Interview - 2 interim Records - monthly
(1) requests to use biology materials (2) voluntarily reads books on biology (3) does biology activities and experiments at home (4) volunteers to bring in materials (5) volunteers to do research	Questionnaires Interviews	Students	Random sample Questionnaire - 10 per class (4 classes) Interview - 3 per class, random stratified sample	Post-program
2. Implementation: How much subject matter content is being covered by individual teachers?	Records	Teachers	all	Records - After observations
	Observation			Observations - 4 times during program
3. Cost: What is the cost of wear and tear on laboratory equipment?	Records	District purchasing department	Records this semester and last semester	Records - post - program
	Questionnaires Interviews	Teachers	all	Quest./Int.--part of measures from #1
4. Research: Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit, is more effective in improving student performance?	Unit retests and Program Posttests	Students	all	As completed
5. Unanticipated Outcome: Have any logistical problems been caused by the field trips?	Records	District accounting office	-	
	Questionnaires Interviews	Teachers	Questionnaire - all 10th grade teachers Interview - random sample of 25% of 10th grade teachers.	post-program

Feedback
Exercise 6
Data Analysis and Interpretation

1. Outcome Question: Do students show a more positive attitude toward and greater interest in biology as evidence by:

   (1) requests to use biology materials

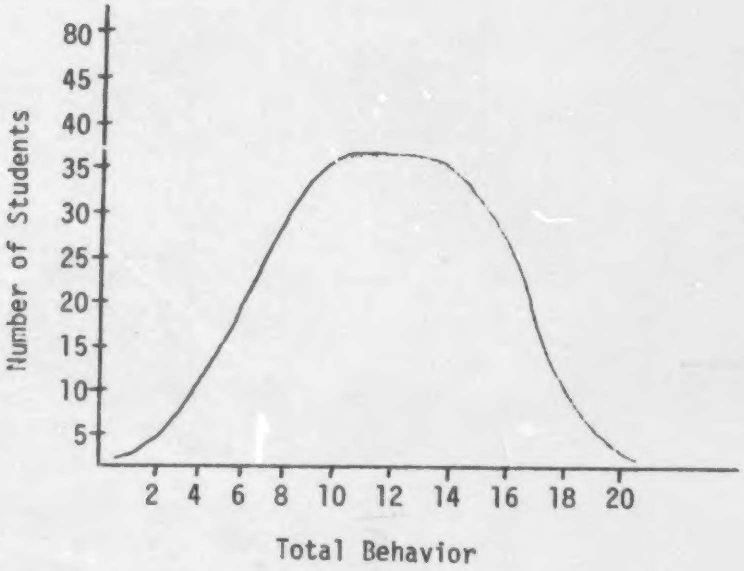
   (2) voluntarily reads books on biology

   (3) does biology activites and experiments at home

   (4) volunteers to bring in materials

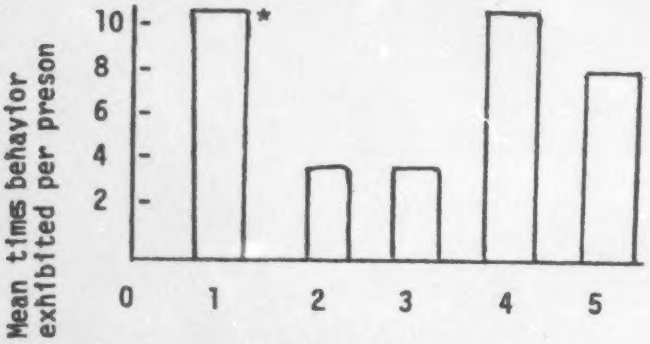
   (5) volunteers to do research

   1. Calculate average number of total behaviors exhibited per student and display in a frequency graph.



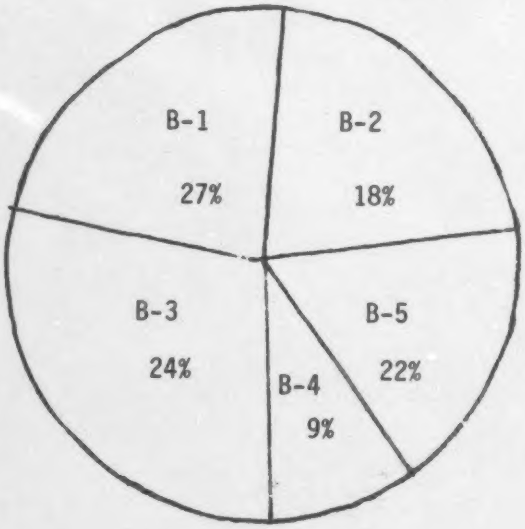
Possible answers:

2. Calculate the average number of times each behavior was exhibited and display in a bar graph.



*hypothetical number

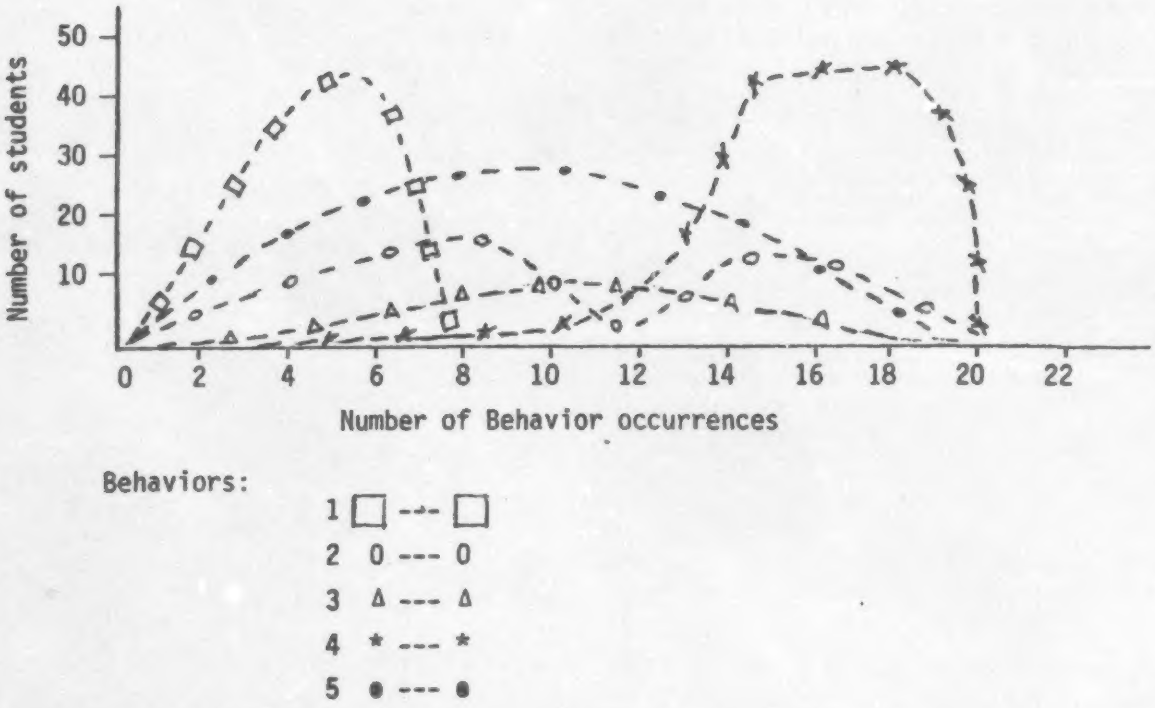
3. Calculate the percentage of students displaying each behavior and display in a pie.



B=Behavior

K-5

4. Calculate number of occurrences of each behavior and display in a graph.



2. What is the cost of wear and tear on laboratory equipment?

The chart shows new program equipment costs are: 1) lower for equipment replacement, 2) higher for equipment purchases. Lower replacement costs may be a function of; students greater interest in and positive attitude toward biology, difference in efficiency and coordination of group of students taking biology this year (apart from interest in program), more supervision by teachers because there is more accountability with new program. Recommendation might be more documentation next semester before any causes can be attributed to lower replacement costs. Higher equipment purchases is probably due to new program requirements and inflation. Therefore, new purchase costs would be expected to be reduced by the next semester. Again a recommendation might be to allow the laboratory session to run another semester without modification, and then document the costs.

3. Research Question: Which remedial technique included in the Living Systems program, the Practice Kit or the Instructional Kit, is more effective in improving student performance?

The table shows an average posttest score of 3 pts (75-72) between the Instructional Kit and the Practice Kit. This difference is statistically significant (t = 3.7). Even though a 3 point increase in test scores is desirable, it is not a large practical significant difference. Given the cost difference, Instructional Kit $1.50 higher per person, it is probably not cost-effective. The planning committee might reasonably suggest using only the Practice Kit for supplementary practice next semester.

4. Unanticipated Outcome Question: Have any logistical problems been caused by the field trips?

Additional Information Desired:

Interviews with administration personnel at the bus company to probe cause of problems.

Conclusions:

Success of field trips and student and teacher reaction to them should be reassessed next semester. Too many problems to draw conclusions this semester.

Recommendations:

Explore possibility of dealing with another bus company. If that is not possible, then before start of next semester meet with bus company representative. Draw up new contract with bus company clearly stating times buses required and assurances that a back-up bus is available in case of breakdowns.

Feedback

Exercise 7

Summary: Formative Evaluation Report

Hamilton Unified School District contracted with the State Evaluation Agency to conduct a formative evaluation of Living Systems, a tenth grade biology program which was implemented at Hamilton High School in September of this year. The purpose of the evaluation was to fully document the program's operation and its impact on students so that any weak program components could be identified and then modified. In order to ensure the collection of relevant information, the evaluation was structured around a series of evaluation questions. These questions dealt with the planned and unplanned effects of one program (outcomes and unanticipated outcomes), the way in which the program was put into operation (implementation), the costs associated with the program (cost), and the effects of varying some program components (research). The measurement devices used to collect information to answer the evaluation questions included unit and end of program tests, questionnaires, interviews, observations. In addition, teacher's records and school accounting records were examined. Information was collected from teachers, students, parents and administrators.

The results of the study by category of evaluation question were as follows:

Outcome: student attitude toward biology and the Living Systems Program

Teacher interviews revealed that all the teachers participating in the program felt students were enthusiastic about the program and about biology. Those students who filled out an anonymous questionnaire at the end of the program indicated enthusiasm with the subject matter and with the Living Systems program.

Implementation: coverage of subject matter by individual teachers.

Observations of teachers' classes at four points during the program and examinations of records teachers kept during the program revealed that the four program teachers taught 85% or more of the lessons as scheduled.

Cost: wear and tear on laboratory equipment

An examination of district purchasing records for the semester prior to the new program and for the first semester of the Living System program's operation showed a lower expenditure in equipment replacement this semester. However, there was greater expenditure on new equipment this semester, a cost teachers felt was associated with installing a new program and would not continue.

Research: effects of two remedial kits.

The relative merit of two remedial kits was compared by analyzing students' retest scores. Both kits appeared to be similarily effective in providing remedial instruction.

Unanticipated outcome: logistical problems with field trips.

District accounting records, questionnaires to 10th graders, and interviews with teachers of other subjects at Hamilton High revealed problems with absences from other classes, conflicts with class scheduling and school activities, and problems arranging bus transportation.